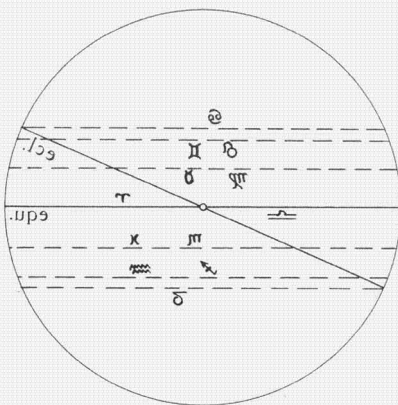# The Mathematical Intelligencer
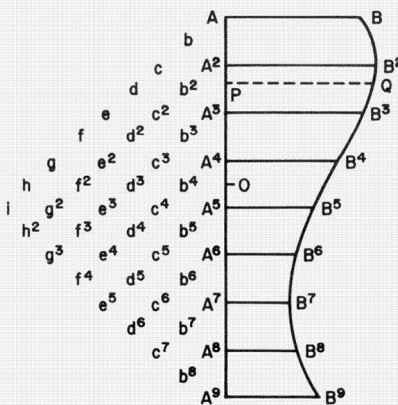
History of Mathematics   A Selection

Studies in the History of Mathematics and Physical Sciences

Editors: M.J. Klein, G.J. Toomer

Volume 1
O. Neugebauer
A History of Ancient Mathematical Astronomy
In 3 parts, not available separately
1975. 619 figures, 9 plates, 1 foldout.
XXIII, VII, V, 1456 pages
Cloth DM 290,--; US $ 145.00
ISBN 3-540-06995-X
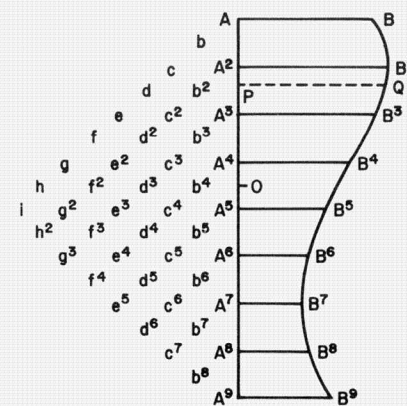
Volume 2
H.H. Goldstine
A History of Numerical Analysis from the 16th through the 19th Century
1977. 23 figures. XIV, 348 pages
Cloth DM 54,--; US $ 27.00
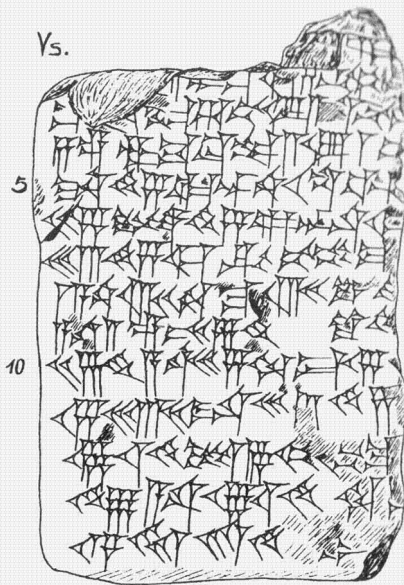ISBN 3-540-90277-5

Volume 3
C.C. Heyde, E. Seneta
I.J. Bienaymé
Statistical Theory Anticipated
1977. 1 figure, 2 tables. XIV, 172 pages
Cloth DM 46,--; US $ 23.00
ISBN 3-540-90261-9

Sources in the History of Mathematics and Physical Sciences

Editors: M.J. Klein, G.J. Toomer

بسم الله الرحمن الرحيم اللهمّ اعمر،
كتاب ذيوقليس في المرايا المحرقة.

Volume 1
Diocles
On Burning Mirrors
The Arabic Translation of the
lost Greek Original.
Edited, with English Translation
and Commentary by G.J. Toomer
1976. 32 figures, 24 plates. IX,
249 pages (64 pages in Arabic,
12 pages in Greek)
Cloth DM 68,--; US $ 34.00
ISBN 3-540-07478-3

O. Neugebauer
Mathematische Keilschrift-Texte
Mathematical Cuneiform Texts.
Edition with Translation and
Commentary in German.
In zwei Bänden. (Reprint der
Erstauflagen Berlin 1935 und 1937)
1973.
Textband: (4) XII, 516 Seiten.
Tafelband: (4) IV, 64 Textseiten,
69 Tafelseiten und VIII, 83 Text-
seiten, 6 Tafelseiten.
Die Originalausgaben sind 1935
(Teil I und II) und 1937 (Teil III,
Ergänzungsheft) als Band 3 der
Reihe "Quellen und Studien zur
Geschichte der Mathematik,
Astronomie und Physik, Abteilung
A" erschienen.
Diese Bände werden nicht einzeln
abgegeben.
Gebunden DM 300,--; US $ 150.000
ISBN 3-540-05617-3

Prices are subject to change
without notice

Springer-Verlag
Berlin
Heidelberg
New York

L. Steen, Saint Olaf College, Department
of Mathematics

---

nts

The diagram on the cover shows an interesting model introduced by
Jürgen Moser in his article on the stability of the solar system (see p. 65).
The oval in the center of the diagram is given, and the point 1 outside the
oval is chosen at random. The motion from the beginning point 1 is by
definition the following: Choose one of the two tangents to the oval
through 1 and follow this tangent from 1
to the point of tangency and continue on
to the point 2, which is defined by the
condition that the point of tangency bi-
sects the segment 12. From 2 draw the
other tangent to the oval and define the
point 3 by the condition that the point of
tangency bisects the segment 23. The
points 4, 5, 6, . . . are defined anal-
ogously. Like the planetary motions, this
motion is easy to describe mathemati-
cally, but conclusions about its global
properties – such as the boundedness
or unboundedness of its orbits – are
extremely difficult to obtain.

The Mathematical Intelligencer

Contents

The diagram on the cover shows an interesting model introduced by Jürgen Moser in his article on the stability of the solar system (see p. 65). The oval in the center of the diagram is given, and the point 1 outside the oval is chosen at random. The motion from the beginning point 1 is by definition the following: Choose one of the two tangents to the oval through 1 and follow this tangent from 1 to the point of tangency and continue on to the point 2, which is defined by the condition that the point of tangency bisects the segment 12. From 2 draw the other tangent to the oval and define the point 3 by the condition that the point of tangency bisects the segment 23. The points 4, 5, 6, . . . are defined analogously. Like the planetary motions, this motion is easy to describe mathematically, but conclusions about its global properties – such as the boundedness or unboundedness of its orbits – are extremely difficult to obtain.

Letters to the Editors

My views on the catastrophe theory controversy must be pretty widely known by now, and I have no wish to repeat them. John Guckenheimer's article (*Mathematical Intelligencer* 1 (1978) 15–20) injects a welcome note of rationality into this dispute. In my opinion he correctly identifies many of its sources; he makes a crucial distinction of *types* of catastrophe theory (though I feel he identifies Zeeman too closely with one particular type); and I strongly support his view that the current discussion is misplaced.

I would like to direct attention to a number of sources bearing on the matters raised by Guckenheimer.

The validity of Zeeman's use of Thom's theorem to infer cusp geometry is addressed by Poston (*On deducing the presence of catastrophes,* Battelle, Geneva 1977). In the context of smooth models the requisite modelling hypotheses are the transversality principle and minimal singularity. Both have ample precedent in the sciences: their use is explicit in Zeeman, *Primary and secondary waves in developmental biology*, Catastrophe theory: Selected papers (1972–1977), Addison-Wesley 1977 pp. 141–233.

Experimental support for the theory of this paper is discussed by Zeeman (op. cit. 267–286). So far Sussmann and Zahler's only gesture towards it (*Nature* 269 (1977)) is to dismiss it. I think it is incumbent upon any critic to *discuss* it.

Recent work applying catastrophe theory to psychology is discussed in a survey by Flay, *Applications of catastrophe theory in psychology*, to appear in *Behavioral Science*, Sept. 1978. While many of the ideas there discussed are very tentative, I do not detect the effect feared by the critics, that social scientists will go overboard for catastrophe models purely because of their impressive mathematical pedigree. On the contrary, the behavioural scientists seem to have a fairly clear judgement both of the likely value of catastrophe models, and the problems in using them. Marc Lewis (*Trend surface analysis of community variables, Psych. Bull.* 1977, 84, 940–949) has made a start on the extremely difficult problem of statistical tests of goodness of fit for nonlinear models. In an abstract for the 1978 APA Annual Convention at Toronto, Lewis states, "Although these results are preliminary, they strongly suggest that catastrophe theory can produce results that are superior to those attained by other methods" (among which he cites curvilinear multiple regression). The use of catastrophe surfaces to model discontinuities in the physical sciences, as Guckenheimer says, is hardly a breakthrough. In psychology, however, it appears to be an important idea.

The idea that "bifurcation" models have analogies with social effects is not only Zeeman's. Haken's book *Synergetics* (Teubner, Stuttgart 1973) includes a very Zeemanesque "pitchfork" bifurcation model of polarization of opinion. Unlike the models of Isnard and Zeeman it is structurally unstable, and this has behavioural consequences.

Guckenheimer remarks of phase transitions " . . . here *the* Catastrophe Theory model is called the Landau model. It is a century old . . . it is also known to give wrong answers." The italics are mine: there is no such thing as "the" catastrophe theory model of a physical system. In our book *Catastrophe theory and its applications*, Pitman, San Francisco and London 1978 (published March 28th) Poston and I note that the Landau theory is just the *most direct* way to apply catastrophe theory to phase transition. (It is still widely used, for its simplicity, in areas where its failure is not acute: for example there is a well-developed analogy between laser action and Landau theory phase transitions, which catastrophe theory formalizes: see chapter 15 of that book.) It is not the *only* way; but other possibilities have yet to be developed very far. In particular there is a tentative connection with renormalization group methods. Inasmuch as a Landau theory singularity organizes behaviour *around* a critical point with some success, and catastrophe theory is the natural language in which to discuss singularities, we may reasonably hope that a more refined use of singularity methods could prove more successful. What we do not yet know is how to achieve this.

There is a close analogy in optics. To use Guckenheimer's terminology: "the" catastrophe theory model of caustics is ray optics, which is at least two thousand years old, and is known to give wrong answers (such as infinite intensity). But catastrophe theory is also a key ingredient in the Arnol'd-Duistermaat development of Maslov's ideas on oscillatory integrals, which is a semiclassical *quantum* optical theory (Poston and Stewart, op. cit. chapter 12). The use of catastrophe theory as a natural language adds power to both classical and semiclassical theories, *and links the two.* A similar development in phase transitions may yet be hoped for: indeed it has been recognized for some time that there is a tantalizing similarity, mathematically and physically, between oscillatory integrals and renormalization group methods, but this has never been successfully formalized.

Ian Stewart
University of Connecticut
Storrs, Connecticut 06268

Editorial

This issue of the Intelligencer contains many different views of the nature of applied mathematics. It is more by coincidence then by design that this has happened. Of course, one might expect the speeches of the three winners of the Birkhoff prize to contain discussions of the subject, but it was unexpected that the three views of applied mathematics would be so completely different and, each in its own way, so illuminating.

The editors were pleasantly surprised, then, to see that Balakrishnan included in his article on the Kalman filter some views – again quite different from those of the Birkhoff laureates – on the theory and practice of applied mathematics. In addition, both the Moser article and the Knorr article raise interesting questions about applied mathematics, especially concerning its role in stimulating work in pure mathematics. Even Hilton's response to Kline touches on the issue of the social importance of applied versus pure mathematics.

We do not necessarily endorse any of these views. In fact, being relatively pure mathematicians, we regard ourselves as observers of a very interesting discussion; we have a vital interest in the outcome but we must be very modest about our competence to judge.

In what way is mathematics valuable in human society, in technology, or in our understanding of the world and the universe? Is the link between mathematics and natural science essential to the health of both, or are there now valuable domains of pure mathematics that are totally divorced from applications outside of mathematics? In a period when academic employment is so hard to find, what subjects should mathematical educators be teaching their students to best serve the interests of these students, of society, and of mathematics? And what, after all, should the term "applied mathematics" mean?

These are questions which we feel should engage the interests of all mathematicians. We hope that what is said about them in this issue will stimulate discussion and that the discussion will continue in later issues of the Intelligencer.

////////////////////////////////////

From: The **Old** *Mathematical Intelligencer, Issue 2 1972*
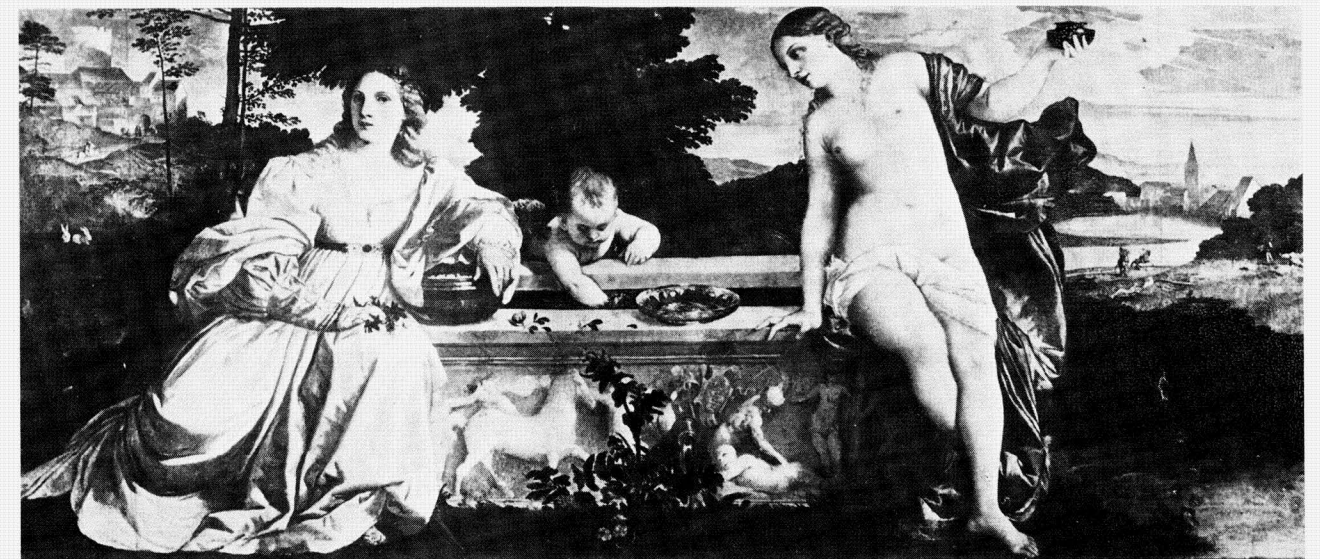
Titian: Sacred & Profane Love / Pure & Applied Mathematics

A reproduction of Titian's famous painting "Sacred & Profane Love" hung in the Göttingen Mathematics Institute during the golden twenties. The meaning, which may not be clear to mathematicians of a younger generation, was gleefully communicated to the Intelligencer by Georg Pólya:

Profane Love, the lady not so well endowed with garments, corresponds to Reine Mathematik, for Sacred Love clearly is "Angewandte Mathematik" ("Gewand" means clothing).

There is, however, a faint hope (for possible applications, of course) because Pure Mathematics is well covered in her more vital parts. In fact, Intelligencer has heard contemporary linguists (not structural) speculating that the terms "Reine" and "Angewandte Mathematik" were inspired by Titian's painting.

This is a pictorial footnote to J. Guckenheimer's article on "The Catastrophe Controversy", Mathematical Intelligencer Volume 1 No. 1.

In a discussion with Thom (see Dynamical Systems Warwick 1974 Springer LN No 468, p 373), Zeeman points out the highly speculative content of Thom's work, and illustrates it by (you have seven guesses) the cusp catastrophe, as follows
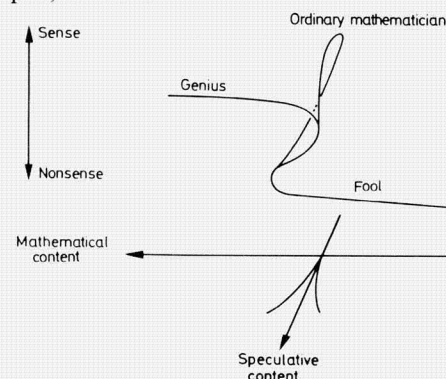
adding: "Thom certainly puts himself out of the ordinary by his courageous speculative ventures, but however close he sails to the edge, he somehow always manages to stay on the upper surface."

Part of the criticism addressed to Zeeman can perhaps be expressed by saying that Zeeman does not always manage to stay on the upper surface in the following "communication catastrophe":
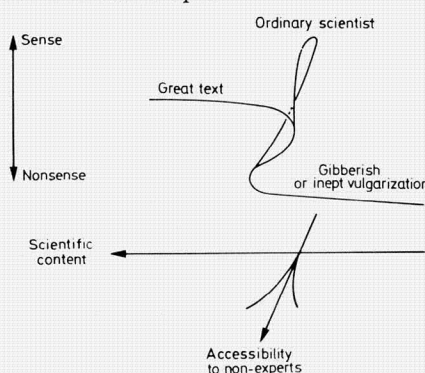
In the same text, Zeeman quotes an analogy of Thom: "Just as when learning to speak, a baby babbles in all the phonemes of all languages in the world, but after listening to his mother's replies, soon learns to babble in only the phonemes of its mother's language, so we mathematicians babble in all the possible branches of mathematics, and ought to listen to mother nature in order to find out which branches of mathematics are natural." I think that Thom proposes that we should use catastrophes as a language, which, it seems to me, implies also that the relation of catastrophes to reality is as complicated as that of the usual language.

We should respect and encourage people like Zeeman who try with daring and imagination to "babble" in this language of catastrophes, even if the result, as some people claim, is not always euphonious with our current scientific language, or even our ideas!

B. Teissier
Research Institute for
Mathematical Sciences
Kyoto University
Kyoto, Japan

Many thanks for sending me the copy of "The New Mathematical Intelligencer". Although I wasn't familiar with the old one, I would like to make a few encouraging remarks.

Nice as the reproduction of the Erlanger Program is, one should really have set Klein's group-theoretic observations in the now established historical context of the theory of Lie groups, as presented for example by Cartan in his lecture at the International Congress of Mathematicians in Oslo. The memorial to Gauss for Gauss Year is of course especially good. But we should perhaps try to beware of acting as if mathematics has only been done in Göttingen.

E. Hölder
Mainz

Is the Solar System Stable?

Jürgen Moser

This article is based on the first of three Pauli lectures, given at the Eidgenössische Technische Hochschule in Zurich in January, 1975. It originally appeared in German in the Neue Zürcher Zeitung of May 14, 1975.

The Stability Problem

The stability problem of classical mechanics, that is, the question as to the stability of the solar system, has fascinated astronomers and mathematicians for centuries. It is simply the question of deciding whether the planetary system in the distant future will keep the same form as it now has or whether after a long time perhaps one or another of the planets might leave the solar system or whether collisions might even lead to a catastrophic change. Since Newton, that is for about 300 years, one has known the laws which govern planetary motion. To a first approximation, the planets move in elliptical orbits in which the sun is located at one of the foci of the ellipse. This is however only a crude approximation to the true motion. The forces between the individual planets cause perturbations so that the form of these elliptical orbits very slowly but steadily changes. The description of these changes, the so-called *secular perturbations* of the elliptical orbits, is the problem that is treated by classical perturbation theory. Now it is conceivable that the relatively weak forces between the planets after a sufficiently long time would so greatly change the present orbits that a planet might be thrown out of the system or that a collision might occur. For example, one can imagine that the eccentricity of a planet might continually increase until its perihelion came so close to the sun that it would meet with misfortune. Although such an eventuality does not agree with our observations over the last millenia, it is something altogether different to *prove mathematically* from the equations of motion that it cannot occur.

As a matter of fact the literature already contains a considerable number of stability proofs. About 100 years after the publication of Newton's *Principia*, Lagrange gave his famous stability proof for the solar system. Further proofs of this type were given by Laplace and Poisson, and one might well ask why the question is again being raised 200 years later. In general one proof is sufficient and the carrying out of several proofs tends to make a critical listener rather suspicious. Actually it is a question here of approximations of varying degrees of accuracy in which the perturbing forces are taken into account only to the first or second powers of the planetary masses. In practice this means that the changes in the elliptical orbits will require a substantial amount of time before they become noticeable. Sommerfeld speaks in his book with F. Klein very tersely of Laplace's "mock proof" *(Scheinbeweis)*, of the stability of the planetary system. How justified these approximations are remains to be seen. When one restricts consideration to a few decades or centuries these stability proofs certainly give the right conclusion, but from this one naturally cannot draw any conclusions about motion in *many millions of years*. Formerly one was primarily interested less in long range predictions than in the practical computation of the positions of the planets, the so-called ephemerides — a question that was of interest already to the Babylonians. Perturbation theory is in fact an outgrowth of the necessity to determine the orbits with ever greater accuracy. This problem can be solved today, but in what is for the theoretician a rather disappointing way. With modern calculating machines, one is now able to compute directly results even more accurately than those provided by perturbation theory. Today the ephemerides of the Nautical Almanac in Washington are computed in this way.

But the mathematical problem only begins here. It is a tried and true technique of mathematics to extract the essential properties of a problem and to idealize it. We deal not with the planets of the solar system, which are after all extended masses, and all kinds of forces are disregarded, such as, for example, the solar wind and relativistic effects. Instead we consider an idealized problem and study n *mass points* which move in 3-dimensional space according to Newton's laws. For the most part one assumes further that n — 1 of these fictious mass points have *very small masses* compared to the remaining one, which plays the role of the sun. Furthermore, we do not ask for the development of the motion for a limited time but *for all eternity*. This is now a purely mathematical problem, the solution of which has a rather limited meaning for the real world, but which entails, by virtue of the demand for a description for all time, very astonishing subtleties. Even this idealized mathematical problem was formulated at least a 100 years ago and is rather vaguely known as the n body problem. In the previous century this problem was of the greatest interest and, as we shall now see, Dirichlet, who today is best known for his monumental works in number theory, and Weierstrass, the function theorist, as well as Poincaré, a universal mathematician, all played essential roles in the treatment of this problem. Thus it is a matter of describing the behavior of the secular perturbations over long time intervals and even for all time. Can changes in the shape and position of the orbits completely alter the configuration of the planetary system? Lagrange proved in connection with his stability proof that these perturbations are subject to

periodic oscillations and thus do not increase without bound. The periods of these oscillations are relatively long, requiring from 5×10^4 to 2×10^6 years. But one must further mention that one is dealing here with an approximation and that in a certain sense this statement can be regarded as a refinement of the age-old description of the planetary orbits by epicycles.

However, what will happen in time intervals of *several million years?* It is a question here of a resonance problem in which the motions of the eight planets play the role of oscillators. Of course, resonance occurs when one deals with a system with a frequency which coincides with one of the eigenfrequencies of the system or an integer multiple of one. The simplest resonance phenomenon is that of pumping a swing. With relatively small forces which are carried out periodically at the frequency of the swing one can increase the amplitide of the swing as high as one wants and can even cause the swing to overturn. In the case of the solar system, such phenomena also play a major role. Indeed, because there is no friction to speak of, any oscillation once established is never damped out. This is the reason why the resonance effects are so subtle for undamped systems in contrast to all everyday physical experiments – or swings. In our solar system there are a great many resonances. For example, it is known that Jupiter and Saturn have a frequency ratio of about 5/2 so that after 5 Saturn years Jupiter has gone through exactly 2 if its years and the forces after this period continue to act in the same direction. This indeed has a strong effect on the orbit of Jupiter, a perturbation whose period to a first approximation is about 900 years.

In reality, however, one must expect such resonances for all rational frequency ratios and even those in which a linear combination of the frequencies with integer coefficients vanishes (commensurable frequencies). This is naturally utterly absurd, for in fact the rational numbe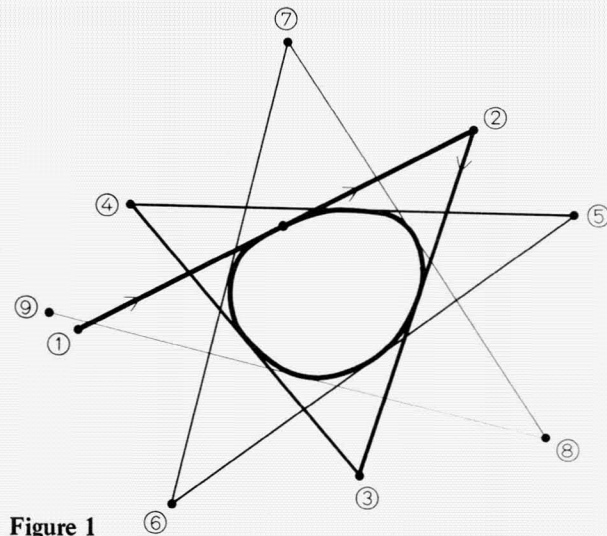rs are dense, and from a physical point of view one can not distinguish between rational and irrational frequencies. On the other hand, the mathematical development definitely requires such a distinction, and we arrive at a paradoxical situation. The equations of motion for the n body problem are very easy to write down but impossible to comprehend intuitively. Therefore it may be useful to describe a very simple geometrical problem that actually contains some of the difficulties of the n body problem and may serve as a crude model for planetary motion (Figure 1). We consider an oval in the plane and define the "orbital motion" in the exterior of the oval as follows. We draw from a point 1 in the exterior one of two tangents to the oval and prolong the tangent to the point 2 which has the same distance from the point of tangency as 1. From 2 we lay out the next tangent to the oval up to point 3, which again has the same distance as 2 from the point of tangency. Continuing in this manner we obtain the "orbit" through the point 1. Can this sequence of points be unbounded? This would be the analogue of the stability problem. Although this problem seems quite elementary it is actually very difficult. One can show that for curves which are smooth enough (admitting 5 derivatives) and have positive curvature the orbits are always bounded, i.e. we have stability.

It is remarkable that in this simple problem the smoothness of the bounding curve should play a role. What happens if corners are admitted? The simplest cases are polygons. Actually the mapping is not continuous in this case but the problem of stability remains clearly meaningful. For general polygons, however, it remains an open question whether the orbits are bounded or not. But there are two special cases which can be fully treated:

1) When the oval degenerates into a 2-gon every orbit goes to infinity along a pair of straight lines (Fig. 2).
2) When the oval is a triangle then all orbits are closed but they have different periods. Points belonging to orbits of the same period form hexagons and triangles which constitute an interesting tesselation of the plane. Points in the hexagons have periods 3, 9, 15, 21, . . . , in general $3(2j - 1)$, $j = 1, 2$ . . . Those of the triangles have period 12, 24, 36, . . . , in general $12j$, $j = 2, 3, . . .$ (Fig. 3). For a square the problem can also easily be handled but even for a general quadrilateral the above question is open.
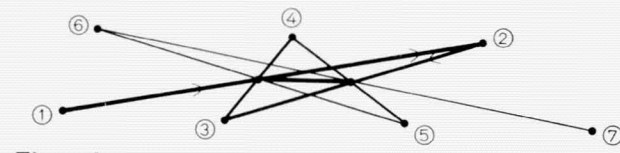
Figure 1
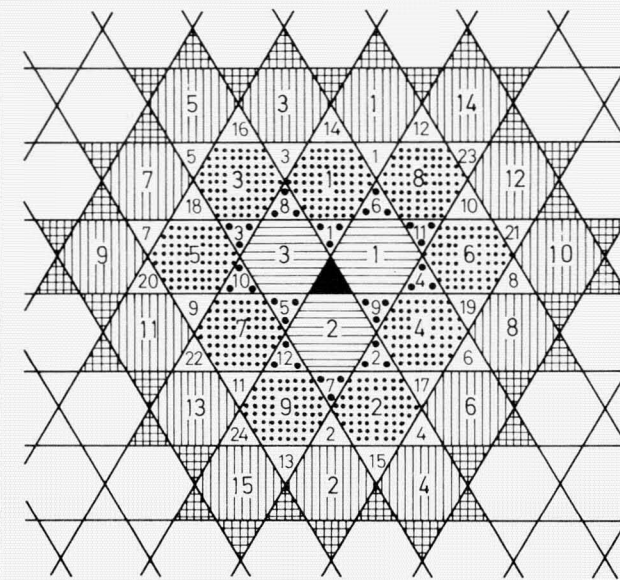
Figure 2

Figure 3

The Prize Question

We return to planetary motion and to the mathematical formulation of the problem and its solution. One attempts to describe the coordinates of the secular perturbations analytically using generalized *Fourier Series.* These are series the terms of which are of the form

$$A_j \cos (j_1 \omega_1 + \ldots + j_s \omega_s) t$$
$$+ B_j \sin (j_1 \omega_1 + \ldots + j_s \omega_s) t$$

with certain frequencies $\omega_1, \ldots, \omega_s$ and combination frequencies $j_1 \omega_1 + \ldots + j_s \omega_s$. This point of view is not far removed from the epicycle theory, but it is mathematically more precise. Such functions are today known as quasi-periodic functions. In fact various mathematicians succeeded, as for example Weierstrass, in obtaining such series developments *formally,* assuming that $\omega_1, \ldots, \omega_s$ are incommensurable (that is their ratios are irrational). But these series may not converge and therefore their usefulness was very much in question. On the other hand, when they converge they describe the small oscillations in the variations of the ellipse. Changes of elements that are thus described remain forever within given bounds.

The mathematical problem can be described as follows: "For an arbitrary system of mass points which attract each other according to Newton's laws, assuming that no two points ever collide, give the coordinates of the individual points for all time as the sum of a uniformly convergent series whose terms are made up of known functions"[1]. This is the word for word translation of a prize question which King Oscar II of Sweden proposed in 1885, that is, 90 years ago. The prize was awarded to H. Poincaré, although he did not in fact solve the problem. His great work actually indicated that such series developments, contrary to all expectations, diverge and thus do not exist. One will be even more surprised to hear that in 1963 an excellent young mathematician in his middle twenties succeeded in solving this problem and in proving the existence of such solutions with complete rigor, at least in the case of the 3 body problem! This mathematician, V. I. Arnold, was a student of Kolmogorov, who a few years before had laid the cornerstone of the proof. More precisely, the breakthrough was based, of course, on the work of many others, and essentially the ideas go all the way back to Poincaré's results.

In the 1940's Siegel solved the first problem of this type. His formulation of the question was, however, more idealized and was not really applicable to mechanical problems. In 1954 Kolmogorov indicated that, for certain mechanical systems, in some sense the *majority* of solutions are quasi-periodic. He indicated a possible method of solution but the actual proof was first provided by Arnold 8 years later, and, in a special case, by the author. In accordance with the modern usage this theory became known by the acronym KAM. The principle result of this theory guaranteed the existence of such quasi-periodic solutions for certain classes of differential equations which included the n body problem. The series developments in question turn out to be convergent for certain choices of the frequencies but are meaningless for other frequencies. This last result was already shown by Poincaré. The orbits which admit such a representation are precisely those for which no resonance occurs. However, since such resonance-free orbits can lie arbitrarily near to the others, it is entirely possible that an arbitrarily small perturbation in the intial values will change a quasi-periodic stable orbit to an unstable one. One can show, however, that the unstable orbits are much rarer, or, as one would say more technically, in phase space have relatively small measure. This means that one is lead to a *new concept of stability* in which the restriction applies only to the *majority* of certain orbits. Whether the relatively rarer unstable exceptional orbits actually exist is still an open problem. We must say at the outset – and it will be shown in what follows – that the weakened concept of stability is very meaningful and satisfactory for the physical applications.

New Applications

But in what does the great progress lie? If the determination of the orbits can be handled very well with computing machines, such a proof seems *superfluous* and at very least *historically too late.* To this one can make the following reply:

1. The stability of undamped systems *for all time* can not in principle be decided by finite calculations and

lies therefore beyond the range of calculating machines.

2. What is more important, however, is that such a result which, by the way, is of the greatest interest in itself from the point of view of mathematics, is also of essential importance to theoretical researches in *statistical mechanics.* The development of statistical mechanics had led to the expectation that most mechanical systems, at least when they are made of sufficiently many particles, are ergodic, that is, after a sufficiently long time their behavior is entirely independent of the initial conditions. This stands, however, in the most striking contrast to stability. In fact, physicists have, beginning with this point of view, in the past century attempted to show that almost all mechanical systems display this unstable behavior provided only that one waits long enough. That this is *not* so for many realistic systems is now, through the work of the last decade, proved once and for all.

3. There is finally a third ground which appears to be more or less coincidental: the mathematical theorems of KAM deal not only with the planetary system but also with general Hamiltonian systems (thus, systems which describe undamped processes of motion) and can therefore be applied to many *other problems.* This is precisely the advantage of a general mathematical formulation. One of these applications is the stability problem of proton accelerators, which since the 1950's have been built in every greater numbers and greater size. The purpose of these machines is to accelerate electrons or protons to extremely high velocities and then to shoot them at a target in order to observe the results of the consequent disintegration, namely, new elementary particles. The greater the energy of the particle is, the more interesting the resulting observations will be. In order to achieve these high velocities, the protons are accelerated in a circular channel more and more until the particles reach a velocity near the velocity of light. These channels, in the case of the proton synchrotron at CERN in Geneva, have a circumference of over 600 meters; and air is pumped out of them in order to create a high vacuum and avoid collisions with gas molecules. A magnetic field is created by a series of magnets, and this field holds the particles in a nearly circular path. This leads to a stability problem because the magnetic field must be constructed in such a way that the protons do not deviate too far from an ideal circular path and thereby lose their energy on the walls of the chamber. In the process, the particles run around the vacuum chamber millions of times.

The question of stability is an essential point in the construction of these accelerators. Although one was at first content to make experiments with calculating
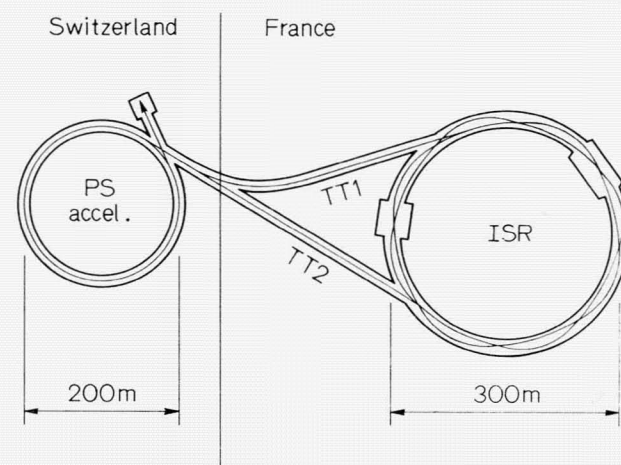

Figure 4 a. Cross-section of the vacuum chamber at the position of the beam inflector, with indication of the stacking process
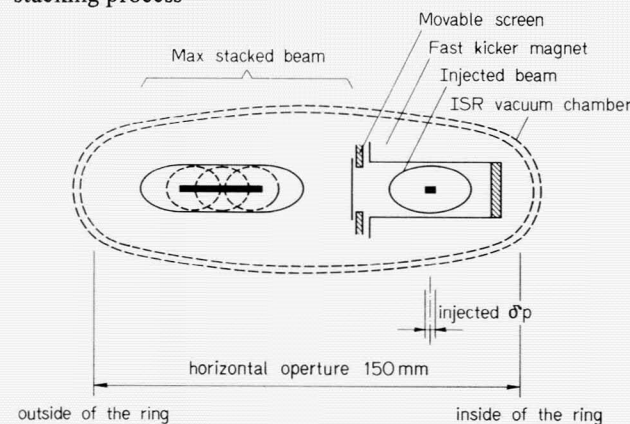
Figure 4 b. Layout of the intersecting storage rings (ISR)

machines, it soon became clear that after a few iterations the unavoidable computational error got out of hand and it became impossible to follow or to predict the paths. One needed *theoretical results* which showed that one could guarantee stability in such a system over a very long time interval, and that is precisely the significance of the theory we are discussing. The latest stage in this development is represented by the so-called *storage rings* of which one has been operating at CERN since 1971 (Figure 4). Roughly speaking, it is a matter of accelerating protons in two circular channels in *opposite directions* and then aiming the beams at one another. In this way the available energy is not only doubled, as one would expect, but, because of relativistic effects, is increased by the square. In order to achieve the motion of protons in opposite directions, one connects such a storage ring (ISR — intersecting storage ring) to the proton synchrotron and introduces bunches of protons alternately in one or the other direction into the storage ring. There they are stored, a process which can last from 3 to 11 hours, until they are made to collide.

The construction of this machine presents *unbelievable technical difficulties* and demands unheard — of precision which would make even Swiss watch manufacturers blanch. There is an obvious comparison to be made here, which is of importance to our stability problem: During the storage process the proton packets must orbit 10^{10} to 10^{11} times around the circular path and in the course of this time be contained within a tunnel that is 16 by 5.2 centimeters. When one equates a circuit of the protons in the storage ring with a year in the astronomical problem, then the above number represents a time which surpasses the age of the earth. That is, one can follow the protons for a longer time, in this analogy, than the solar system has existed in its present form. In addition to this, the experimental physicist or technician can alter the conditions and the parameters at will. We discuss this example here because it requires stability over time intervals which exceed anything that was dreamed of in astronomy 100 years ago and therefore in a certain sense justifies the idealized stability question concerning infinite time intervals, if indeed such justification were necessary. When one applies the results of KAM theory in this situation one finds that the majority of accelerated protons are conserved within the circle of the storage rings but that the relatively rare exceptional orbits lead to a slow and very slight loss in the proton rays. Such losses are in any case unavoidable and are also observed. Whether these exceptional orbits can be considered to be responsible for this loss must be regarded as still unresolved in view of the fact that many additional forces and effects which affect the particles and may deflect them have been neglected. Such applications can provide a strong stimulus to mathematical research. One must certainly ask how it is that this age-old stability problem, just when it is losing its interest for astronomers, has suddenly been solved. One can well hypothesize that the development of proton accelerators has influenced the rebirth of interest in this question.

Stability of Periodic Orbits

We wish to describe another resonance phenomenon, which enters both in astronomy and in high energy physics. It is the question of the stability of *periodic orbits,* that is, of those solutions which after a certain time return to their initial configuration. One asks for conditions under which all orbits whose initial conditions lie near the periodic orbit remain for all time near the periodic orbit. Such orbits are called stable. Only such stable orbits are normally observable. The best example is the circular orbit in a storage ring described above. Small perturbations should not lead to large deviations. In order to determine whether these circular orbits are stable, one must use the so-called betatron frequencies ω_r, ω_z and the orbital frequency ω_o which belong to the oscillations of the linearized system. The theory shows that in general nonlinear resonance or instability will occur when the frequencies satisfy a relation

$$n\omega_r + m\omega_z = p\omega_o$$

with whole numbers n, m, p for which $|n| + |m| \leqslant 4$; such relations with $|n| + |m| > 4$, on the other hand, are harmless. Experiments show that in fact in the first case a loss in the beam is observed but in the second case this loss is negligible. Loosely speaking, resonances of order less than or equal to 4 are in general dangerous, whereas those of order greater than 4 are harmless.

An analogous phenomenon occurs in astronomy. As is well known, in addition to the major planets there are *many of thousands of asteroids* circling the sun; their orbits are primarily between those of Mars and Jupiter. Their masses are minescule and therefore have no influence on the planets. On the other hand, the asteroids are very substantially perturbed by Jupiter. Evidence for

Fig. 5. The number of asteroids as a function of the semi-major axis a. The a-values corresponding to certain fractions of Jupiter's period are marked below. Some of these 'resonances' have produced gaps in the asteroid distribution.

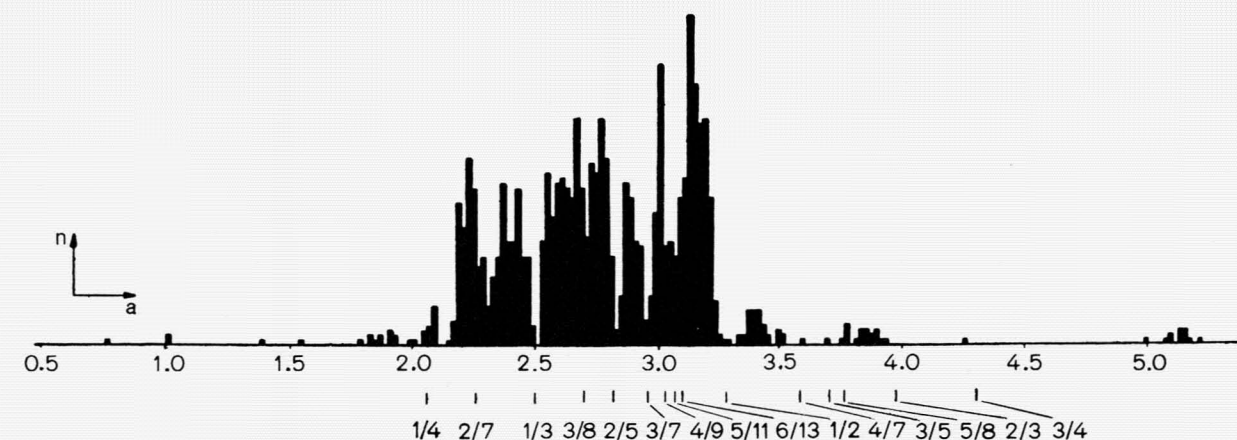this is an observation due to Kirkwood. He remarked that the frequencies of the asteroids are not uniformly distributed over an interval but that there are certain gaps, the so-called Kirkwood gaps, to be observed (Figure 5). One can consider this situation to be analogous to the gaps in the rings of Saturn, which in fact present a similar phenomenon. If the mean motion of the asteroids is denoted by ω_a and that of Jupiter by ω_j then the most pronounced gaps are given by the formula

$$\frac{\omega_j}{\omega_a} = \frac{n}{m} \qquad |n - m| = 1,2,3,4$$

and this means that it is a matter of resonances of order $\leqslant 4$. It remains to characterize the periodic orbits whose stability corresponds to the above conditions. One imagines Jupiter on an exactly circular orbit and lets the asteroids move on a nearly circular orbit in the same plane in such a way that the configuration, that is, the triangle formed by the sun, Jupiter, and asteroid returns to its original position after a certain period of time. Such periodic orbits were already derived by Poincaré. The orbits for which the resonance given above does not occur are stable, so that the explanation is obvious: the gaps correspond to unstable orbits. Although these are only crude approximations to the actual situation, they nonetheless successfully reflect the phenomenom of gaps. The mathematical explanation of this phenomenon is given rigorously by the KAM theory, although an essential idea can already be found in the work of Birkhoff, who continued Poincaré's work.

Historical Remarks

The following historical sketch illuminates the stability problem and its very dramatic development. These remarks were stimulated by the fortunate circumstance that the letters of Weierstrass to Sonya Kowalevsky were published a short time ago. These letters contain much interesting material on our subject which otherwise is very little known, even to mathematicians. Weierstrass played an absolutely central role in the mathematical life of the second half of the 19th century, and mathematicians from all over the world came to Berlin to hear his lectures. His principal interest and his life's work was function theory, but he also had a serious interest in astronomy and gave a seminar on perturbation theory in astronomy in 1880/1881. His ideas on this subject, and above all on the stability problem, were described in several letters to Sonya Kowalevsky. In view of the fact that Weierstrass published his results only very reluctantly and only after extensive and thorough — going revision, these informal communications are particularly valuable.

Sonya Kowalevsky came to Berlin from Russia to study mathematics in a truly adventurous way as a 20 year old student. It was certainly not customary to have female students there, and she was barred from lectures, as a result of which Weierstrass gave her private instruction. From this developed a close friendship which endured until the end of Kowalevsky's life. Moreover, Kowalevsky developed into a well-known and celebrated mathematician, and was professor of mathematics in Stockholm, although she unfortunately reached this position only two years before her untimely death at the age of 41. In addition to personal communications, Weierstrass' letters to Sonya Kowalevsky contain a large number of mathematical ideas and proposals to his student giving a valuable insight into his manner of thinking.[2] However, one also finds very specific indications, as for example in the letter of 15 August 1878, that already at that time he was in possession of the formal series developments for quasi-periodic solutions of the n body problem, and was occupied with the question of their convergence. After this, there can remain no doubt that Weierstrass was on the track of exactly the same problem which has now been finally solved.

Why was Weierstrass so confident that his series representations actually converged? This is also known. Dirichlet, who was Gauss' successor in Göttingen, had already told his student Kronecker in the year 1858 that he had discovered an entirely new general method for dealing with and solving problems of mechanics. Dirichlet died the following year without leaving behind anything written about his discoveries, but Kronecker communicated Dirichlet's remarks to the mathematical world, which sought to recover this lost idea. Nowadays one connects the name of Dirichlet principally with number theory, which was indeed his main interest, while his works in mathematical physics are less well known. These include the foundations of the theory of Fourier series, stability figures of rotating fluids, hydrodynamical works, stability criteria for equilibrium, and others. Because Dirichlet's publications were known for the absolute rigor of their methods and their proofs, there was little doubt that Dirichlet's remark should be taken seriously, and Weierstrass was particularly interested to clarify this problem and to recover this treasure. When in 1885 a prize was established at Mittag-Leffler's instigation for an important mathematical discovery, Weierstrass proposed precisely this problem as one of the prize questions, as was already mentioned. The committee consisted of Weierstrass, Hermite, and Mittag-Leffler. This lead, then, to the famous work of Poincaré of over 200 pages which had a great effect on the later development of the subject. But for Weierstrass, who expressed his greatest admiration for this work of Poincaré, this was nonetheless a disillusionment, for Poincaré showed that the series developments of perturbation theory *in general diverge* and thereby Weierstrass' hopes appeared to be destroyed. By the way, these divergence phenomena bothered Poincaré

very little and he overcame them in a very bold manner. The asymptotic series developments that one uses in the theory of flows and other applied subjects go back to Poincaré.s ideas, as does the use of divergent series in numerical calculations. Weierstrass, by contrast, pursued the convergence question mercilessly and found that, while Poincaré's deductions were quite correct, they did not in fact prove the divergence of the series in question. The existence theorems for quasi-periodic solutions, which we now know, say precisely that such series do converge for *certain frequencies,* and that was precisely Weierstrass' point. Thus, 70 years later, this question of Weierstrass can finally be given a positive answer. Naturally it is no longer possible to determine whether the new results actually coincide with Dirichlet's attempts or are related to them.

What has been said here should not give the false impression that mathematics is guided solely by such practical applications or that the justification of its existence is to be found in the solution of such problems. Rather it is the vigorous interaction of various areas of study that always leads to new concepts. Is the solar system stable? Properly speaking, the answer is still unknown, and yet this question can lead to very deep results which probably are more important than the answer to the original question.

1. Es sollen für ein beliebiges System materieller Punkte, die einander nach dem Newton'schen Gesetze anziehen, unter der Annahme, dass niemals ein Zusammentreffen zweier Punkte stattfinde, die Coordinaten jedes einzelnen Punktes in unendliche, aus bekannten Functionen der Zeit zusammengesetzte und für einen Zeitraum von unbegrenzter Dauer gleichmässig convergirende Reihen entwickelt werden.

2. *Briefe von Karl Weierstrass an Sofie Kowalewskaya 1871–1891,* "Nauka", Moscow 1973

Courant Institute of
Mathematical Sciences
New York University
New York, USA

Colouring Problems

W. T. Tutte

The Four Colour Problem has been solved by K. Appel, W. Haken and J. Koch [1, 2, 3, 10, 14]. But what about the other mathematicians who have been working on the problem? I imagine one of them outgribing in despair, crying "What shall I do now?" To which the proper answer is "Be of good cheer. You can continue in the same general line of research. You can study the Hajós and Hadwiger Conjectures. You can attack the problem of 5-flows, and you can try to classify the tangential 2-blocks."

For the Four Colour Problem is just one member, one special case, of a great association. It was singled out for mathematical attack because it seemed likely to be the easiest member. But now the time has come to confront the other members of the family.

Let us recall what the Four Colour Theorem asserts. We imagine the entire surface of a spherical planet to be partitioned among various countries. The territory of each country is to form one connected piece. A map-maker wishes to assign a colour to each country so that no two countries with a common line of frontier are given the same colour. He is however willing to give the same colour to two countries meeting at only one point, or at only two points, or at only a finite number of points. The Theorem assures him that he can complete his map using at most four distinct colours.

In discussions of the map we may speak of "vertices", that is points at which three or more countries meet, and of "edges", that is lines of frontier extending from one vertex to another without passing through a third. When the map is considered as a mathematical structure the vertices and edges assume equal importance with the countries.

It is usual to make some simplifying assumptions. For example we will suppose that each country has common frontier lines with at least two neighbours. Such assumptions are justified by proving that if the Four Colour Theorem holds for all maps satisfying them it is easily extended to the others.

The vertices and edges of a map provide an example of the structure known as a "graph". A graph has things called "vertices", usually described as geometrical points and drawn as dots. It also has objects called "edges", each edge being associated with two vertices called its "ends." An edge is usually drawn as an arc joining its ends. In all "drawings" of a graph in a surface, or in another graph, we assume that there are no crossings. This means that no edge is to pass through a vertex which is not one of its ends, and no two edges are to meet except at a common end. As a rule it is best to think of graphs as drawn in 3-dimensional space, so that crossings can easily be avoided. Graphs that can be drawn without crossings on the sphere, like the graphs of maps, are called "planar".

The most general of the definitions in use allows a graph to have an edge with two coincident ends, one joining a vertex to itself. But it is usual to exclude such loops in the theory of colourings.

The "valency" of a vertex of a graph is the number of edges having it as an end. In a general graph we may find vertices of valency 0, 1 or 2, but we do not expect such things in political maps.

Some colouring problems are concerned with (loopless) graphs. An n — colouring of a graph G is defined as a colouring of its vertices in n or fewer distinct colours so that no two vertices of the same colour are joined by an edge. There are some interesting conjectures concerned with the conditions under which a given graph can have an n — colouring.

The colourings of a map M can be represented as colourings of a graph. For let us mark one point in each country, to be called the "capital". Across each edge let us draw an arc called a "railroad" joining the capitals of the two adjacent countries. The capitals and railroads are the vertices and edges respectively of a planar graph G, for we can easily arrange to draw them without crossings. Obviously G has a 4 — colouring on its vertices if and only if M has a 4 — colouring of its countries. We call G the "dual graph" of M.

Using dual graphs we can show that the Four Colour Theorem is equivalent to the proposition that every loopless planar graph has a 4-colouring on its vertices. Indeed it is in this form that the Four Colour Theorem is now usually stated. A change of fashion from countries to vertices took place in the Nineteen Forties.

We might recommend our despondent tetrachromatologist to turn his attention from spherical maps to general graphs. We might begin by telling him about Brooks' Theorem [5].

Brooks' Theorem asserts that if k is an integer not less than 3, and if G is a connected graph in which the valency of no vertex exceeds k, then either G has a k — colouring or G is a complete $(k + 1)$ - graph.

By a complete n — graph is meant a graph with exactly n vertices, each two being joined by a single edge. In a complete $(k + 1)$ - graph the valency of each vertex is k, but there is no k — colouring.

At the critical stage of Brooks' proof he has all but one of his vertices properly coloured in k colours. The remaining vertex has an extra colour, let us call it "pink". Brooks pushes pink around the graph from one vertex to an adjacent one, adjusting the legitimate colours as he goes. At last he gets it to a vertex now joined to at most $k - 1$ legitimate colours. He can now replace pink by one of the k.

Brooks' Theorem is associated with a fascinating unproved proposition called Hajós' Conjecture. It says that if a graph G cannot be coloured in k colours then a complete $(k + 1)$ - graph K can be drawn in G. This time there are no restrictions on the valencies [17].

When we say that K is drawn in G we mean that $k + 1$ vertices of G are specified as the vertices of K. We mean also that the edges of K are drawn between them along the edges of G, and perhaps through other vertices of G. Thus each edge of K appears as a string of one or more edges of G. There are of course to be no crossings.

The Four Colour Theorem is merely part of the special case k = 4 of Hajós' Conjecture. For if some planar graph G cannot be 4 — coloured, then by Hajós' Conjecture we can draw in it a complete 5 — graph K. But this is impossible; the complete 5 — graph is well-known as non-planar.

We might try to prove the Conjecture by adapting the proof of Brooks' Theorem. But the writer does not recommend this. Pushing unwanted colours around a general graph is likely to induce a state of frustration resembling that of a Kgovjnian Commander-in-Chief [6].

Hajós suggested another method of attack [13]. It is based on a result known as Hajós' Theorem. This asserts that any graph with no k-colouring has a subgraph that can be formed from complete $(k + 1)$ - graphs by repeated application of two simple constructions. In one of these two vertices of a non-colourable graph, not joined by an edge, are identified. The second is shown in Figure 1. We start with two disjoint non-colourable graphs G and H. An edge ab is deleted from G, and an edge xy from H. Then a and x are identified, and a new edge is brought in to join b and y. All graphs formed by such constructions from complete $(k + 1)$ - graphs are clearly non-colourable. And Hajós proves a converse of this result.
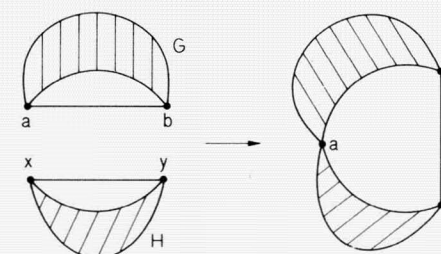
Figure 1. A Hajós join

The Hadwiger Conjecture is rather like the Hajós Conjecture, but it seems to assert rather less. It says that if a general graph G cannot be coloured in k colours then it contains $k + 1$ disjoint connected subgraphs such that each two are joined by an edge of G. It has been shown that the case k = 4 of Hadwiger's Conjecture is precisely equivalent to the Four Colour Theorem; each proposition can be deduced from the other [17, 22, 23].

It is interesting to speculate on the relation between the Hajós and Hadwiger Conjectures. It is easy to prove the second if the first is assumed. But would it be possible to deduce the first from the second without actually proving either? Even the settlement of this logical point would seem like real progress in the theory of graph colourings.

Let us go back to the Four Colour Problem and discuss those maps on the sphere that are called triangulations. In these each country is a triangle; its border is made up of three edges and three vertices.

The face-colouring, that is country-colouring, problem for triangulations is easy. We have only to apply Brooks' Theorem to the dual graph. Excluding the case of four triangular countries mutually adjacent we find that three colours suffice.

The Four Colour Theorem in its usual modern form tells us that the graph of a triangulation can be vertex-coloured in four colours. Moreover it has long been known that if the Theorem holds for all triangulations it must be true for all loopless planar graphs. For, any such graph, drawn on the sphere, can be converted into a triangulation by adding more edges, and perhaps more vertices. Accordingly most papers nowadays on the Four Colour Theorem, including those describing the recent proof, are concerned only with triangulations.

If we are given a 4-colouring of the vertices of a triangulation T we can derive from it a 3-colouring of the edges, in colours α, β and γ, so that each country has one edge of each colour. The rule is as follows. Let the four vertex-colours be 1, 2, 3 and 4. Then an edge is coloured α if its ends have colours 1 and 2, or 3 and 4. It is coloured β if its ends have colours 1 and 3, or 2 and 4. It is coloured γ in the remaining cases.

It is convenient to use here the simple algebraic structure known as the 4 — group. This has exactly four elements 0, α, β and γ. A commutative and associative addition is defined for them. The rules are that $0 + x = x$ and $x + x = 0$ for each element x, and that the sum of any two of the non-zero elements α, β and γ is the third. If the four vertex — colours of our triangulation are taken to be 0, α, β and γ then we can define the colour of an edge as the sum of the colours of its ends. There results an edge-colouring of T in the three non-zero colours α, β and γ.

Let us transfer such an edge-colouring to the dual graph G of T. The graph G is connected, loopless and planar. It is also "cubic", that is, exactly three edges of G meet at each vertex. It can be shown to have no "isthmus", that is, no edge whose deletion would disconnect the graph.

A "Tait colouring" of a cubic graph is a colouring of its edges in three colours α, β and γ so that each vertex is incident with one edge of each colour. A proposition

which has sometimes been called "Tait's Conjecture" and sometimes "Tait's Theorem" says that any connected planar cubic graph without an isthmus has a Tait colouring. Such graphs are simply the dual graphs of the triangulations, and their Tait colourings correspond to the edge-colourings of the triangulations. Accordingly Tait's Theorem can be deduced from the Four Colour Theorem. Actually the two theorems are equivalent.

There is a theory of the Tait colourings of cubic graphs that are not necessarily planar. In view of the equivalence just mentioned we can say that it is concerned with another generalization of the Four Colour Problem.

In this theory we usually ignore cubic graphs having an isthmus. All that need be said about them is that they have no Tait colourings; the proof of this is very simple. We also take little heed of cubic graphs that can be decomposed, by the deletion of either 2 or 3 edges, into two smaller graphs each having a circuit. Their Tait colourings depend, in a trivial way, on those of simpler cubic graphs. Those connected cubic graphs that now remain are said to be "cyclically 4 — connected".

At present the theory is mainly concerned with the search for cyclically 4 — connected cubic graphs with no Tait colourings. Such graphs were called "Snarks" recently in the Mathematical Games Section of Scientific American. The simplest and best-known Snark is the Petersen graph, shown in Figure 2.
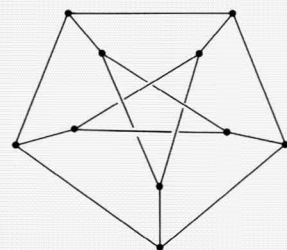
Figure 2. The Petersen graph

A cyclically 4 — connected non-Snark is shown in Figure 3, complete with a Tait colouring.
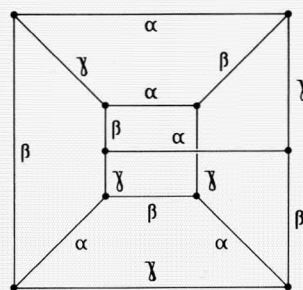
Figure 3. A Tait-coloured graph

A few Snarks other than the Petersen graph were found by D. Blanusa, B. Descartes and G. Szekeres [4, 9, 18]. But Snark — hunting became really productive only with the publication of a paper by R. Isaacs [15].

Here we find techniques for snaring Snarks in infinite families.

We might hope for a complete classification of Snarks, describing each particular batch. Less ambitiously we might try only to prove a conjecture suggested by those of Hajós and Hadwiger, that in any Snark we can draw a Petersen graph. This Conjecture implies the Four Colour Theorem since the Petersen graph is non-planar. It is therefore a true generalization of the Theorem, and it is a challenge to all newly liberated tetrachromatologists.

Map colourings and Tait colourings are curiously related, via some theorems of the writer [19], to "n-flows" in a graph. Imagine some fluid circulating in a graph G, flowing along its edges. The magnitude of the current in each edge is to be an integer. At each vertex, naturally, the sum of the magnitudes of the currents flowing in is equal to the sum of the magnitudes of the currents flowing out. Such a circulation is called an n-flow in G if two conditions are satisfied. First the magnitude of each edge — current must be less than the positive integer n, and secondly no edge-current is to be zero.

The theorems just mentioned show that in cubic graphs 4 — flows and Tait colourings are interconvertible. A Snark therefore can be defined as a cyclically 4 — connected cubic graph with no 4 — flow. It would be interesting to know if there are any Snarks without 5 — flows. A Snark-hunter encountering such a beast would undoubtedly find that some of his pet conjectures had softly and suddenly vanished away. May we not therefore refer to such a Snark as a Boojum? [7].

It is found that if the countries of a plane or spherical map can be 4 — coloured then the corresponding graph has a 4 — flow. Similarly if the countries can be 5 — coloured then the graph has a 5 — flow, and so on. An algebraic topologist, working with coefficients in the integers mod 5, can think of the 5 — colouring as a 2 — chain, and of the 5 — flow as its boundary. A rather well known unproved proposition asserts that every graph without an isthmus has a 5 — flow. In particular the Conjecture would have us believe that there are no such things as Boojums.

Let us note well that the 5 — flow Conjecture generalizes not the terrible Theorem of the Four Colours but the tractable and long-established Five Colour Theorem. Then how does it contrive to be so difficult? There are corresponding unproved conjectures about 6 — flows and 7 — flows. But F. Jaeger has shown that every graph without an isthmus has an 8 — flow [16].

We conclude with a few remarks about the geometrical approach to the Four Colour Problem. This was first explored by O. Veblen [21] in 1912. We consider finite projective geometries in which the points are represented by vectors over the field of residues mod 2. For some purposes a graph G can be represented by a set of points in such a space. The points of this set S correspond to the edges of G. Commonly it is arranged for a subset of S to be linearly dependent if and only if it includes all the edges of some circuit of G. If the projective space is

n-dimensional it is then found that 4 — colourings of G are represented by those (n — 2)-dimensional subspaces that contain no point of S.

Let P be the n-dimensional space, and Q one of its (n — 2)-dimensional subspaces not meeting S. Then we can project P from Q onto a line, a 1 — dimensional space. The points of this line are the vectors (1, 0), (0, 1) and (1, 1) over GF(2), and these can be identified with the non-zero elements α, β and γ of the 4 — group. Each edge is projected onto one of these three "colours". The colours of the edges sum to zero round every circuit, by the common rule of linear dependence. From this fact we can deduce the existence of a 4 — colouring of G in elements of the 4 — group such that the projection-colour of any edge is the sum of the colours of its ends.

Having related 4 — colourings of G to subspaces of P we can now state a wide generalization of the Four Colour Problem. The new problem is that of classifying those sets of points in P that meet every (n — 2)-space in P. A weakened form of the problem presents us with another Hajós-style conjecture.

Let us describe a point-set S in P that meets every (n — 2)-space as a "2-block", whether or not it represents a graph. Let Z be any subset of such a 2-block S. Let T be an (n — 2)-space in P containing all the points of Z, and therefore all the points of S linearly dependent on Z. If T contains no other point of S we call it a "tangent" of Z. The 2-block S is said to be "tangential" if every non-null proper subset Z of S has a tangent. At present the main interest in this geometrical research lies in the quest for tangential 2-blocks.

The importance of tangential 2-blocks is that all other 2-blocks can be regarded as derivatives of them. For consider a 2-block S for which some non-null proper subset Z has no tangent. It can be shown that by projecting P from Z onto a suitable space of lower dimension we convert S into a 2-block in that space. Repetition of this procedure must lead us eventually to a tangential 2-block, from which the original 2-block is in a sense derivable.

So far only three tangential 2-blocks have been found. There is the Fano block, which is the finite projective plane of 7 points. There is the Desargues block, given by the 10 points of a 3-dimensional Desargues configuration. It can be regarded as representing the complete 5 — graph. Finally there is the 5-dimensional Petersen block, which represents the Petersen graph in a dual manner, with cut-sets replacing circuits in the definition of linear dependence.

The conjecture mentioned above asserts that the only tangential 2-blocks are the Fano, Desargues and Petersen blocks. It implies as special cases the Four Colour Theorem and the proposition that a Petersen graph can be drawn in every Snark.

The writer has shown that there are no other tangential 2-blocks of up to 5 dimensions [20]. B. T.

Datta has extended the search to spaces of 6 and 7 dimensions. He reports that no new tangential 2-blocks exist in these geometries.[8]

I hope I have said enough to indicate that the work of Appel, Haken and Koch should be regarded as a beginning and not as an end. The Four Colour Theorem is the tip of the iceberg, the thin end of the wedge and the first cuckoo of Spring. And to thee, discouraged Mapman, contemplating thine abortive proofs, one final word. "Impenetrability! That's what I say!" [11].

REFERENCES

1. K. Appel and W. Haken, *Every planar map is four colorable. Part I: discharging.* Illinois J. of Math. 21 (1977), 429–490.

2. K. Appel, W. Haken and J. Koch, *Every planar map is four colourable. Part II: reducibility.* Illinois J of Math. 21 (1977), 491–567.

3. F. Bernhart, *A digest of the Four Color Theorem.* J. Graph Theory, 1 (1977), 207–225.

4. D. Blanuša, *Problem Cetiriju Boja,* Hrvatsko Prirodoslovno Društvo Glasnik, Mat-Fiz Astr., Ser. II, (1946), 31–42. (Croatian, French summary).

5. R. L. Brooks, *On colouring the nodes of a network.* Proc. Cambridge Phil. Soc., 37 (1941) 194–197.

6. Lewis Carroll, *A Tangled Tale,* Knot VIII.

7. Lewis Carroll, *The Hunting of the Snark,* Fit the Eighth.

8. B. T. Datta, *Non-existence of six-dimensional tangential 2-blocks.* J. Combinatorial Theory 21 (1976), 171–193.

9. Blanche Descartes, *Network colourings,* Math. Gazette 32 (1948) 67–69.

10. Blanche Descartes, *On some recent progress in combinatorics.* J. Graph Theory 1 (1977), 192.

11. H. Dumpty. As reported by Lewis Carroll in *Through the Looking-glass,* Chapter VI.

12. H. Hadwiger. *Ungelöste Probleme.* Element Math. 13 (1958), 127–128.

13. G. Hajos. *Über eine Konstruktion nicht n-farbarer Graphen.* Wiss. Zeitschr. Martin Luther Univ. Halle-Wittenberg A10 (1961) 116–117.

14. W. Haken, *An attempt to understand the Four Color Problem,* J. Graph Theory 1 (1977), 193–206.

15. R. Isaacs, *Infinite families of non-trivial trivalent graphs which are not Tait colorable.* Amer. Math. Monthly, 82 (1975), 221–239.

16. F. Jaeger, *On nowhere-zero flows in multigraphs.* Proc. Fifth British Combinatorial Conference, 373–378. (Utilitas Mathematica, Winnipeg, 1976).

17. O. Ore, *The Four Color Problem,* Academic Press, New York 1967.

18. G. Szekeres, *Polyhedral decompositions of cubic graphs.* Bull Austral. Math. Soc., 8 (1973) 367–387.

19. W. T. Tutte, *On the imbedding of linear graphs in surfaces.* Proc. London Math. Soc., Ser. 2, 51 (1949), 474–483.

20. W. T. Tutte, *On the algebraic theory of graph colorings.* J. Combinatorial Theory 1 (1966), 15–50.

21. O. Veblen, *An application of modular equations in Analysis Situs.* Ann. of Math 14 (1912), 86–94.

22. K. Wagner. *Bemerkung zu Hadwigers Vermutung.* Math. Ann., 141 (1960), 433–451.

23. K. Wagner, *Beweiss einer Abschwächung der Hadwiger-Vermutung.* Math. Ann., 153 (1964), 139–141.

University of Waterloo
Waterloo, Ontario, Canada

Teaching and Research: A False Dichotomy

A Response to Morris Kline

P. J. Hilton

The following response to Morris Kline's book **Why the Professor Can't Teach: Mathematics and the Dilemma of University Education** *was prepared at the invitation of the editors. Excerpts from the book appeared in Issue Number 1.*

Morris Kline has done it again! Five years ago he published his polemic *Why Johnny Can't Add,* (St. Martin's Press, 1973) in which he stigmatized the community of research mathematicians for their culpable responsibility for those distortions and errors of pre-college education which he subsumed under the general head of 'the new mathematics'. Now he has published *Why the Professor Can't Teach,* (St. Martin's Press, 1978), in which he fires yet another broadside at the research establishment, this time for its total inability to teach at the undergraduate level.[1] Actually in a sense he is now attacking the whole university system for it is his claim that it is through the preconceptions and misconceptions of those who administer this system that research has been given priority over teaching to the dire detriment of the latter; and that thereby the undergraduate students of our institutions of higher education have been grossly abused. Kline is prone to personalize the universities who, he alleges, carry out such very human activities as *asserting, maintaining,* and *solving problems.* In particular, universities 'maintain' that researchers are "ipso facto good teachers" and 'assert' that "to be a good teacher, one must be a good researcher." It is hard to confirm or deny that universities make such statements, but I have never heard such a statement made by any single human being, nor of course could such a statement, even if made by so impressive a collective organism as that which seems to communicate its views to Morris Kline, conceivably be justified. The most that any reasonable person would maintain – and I would myself maintain this – is that there exists a positive correlation between successful research activity and good teaching. My reasons for expecting this are the following: I

understand, by 'good teaching', teaching which induces both understanding and interest in the student. These consequences of good teaching are to be observed not so much in the student's immediate reaction to the learning experience, but in his subsequent activities. If the student chooses voluntarily to take further mathematics courses at the university, I would say that is evidence of good teaching. If, in his subsequent life and career, he chooses to place himself in situations in which he is required to use mathematics and enjoys those activities, then he has been taught effectively and well. I do not accept the student's success in tests which take place during and immediately following the termination of a given course as evidence relevant to the question whether the teaching of that course was good. Given the criteria I have proposed, it follows that a highly significant factor conducive to good teaching is the interest of the teacher in his subject. Insofar as mathematics is concerned, that interest is predominantly displayed in research activity. Moreover, research activity tends to reinforce the mathematician's interest in his subject. Thus, while the criterion of commitment to active research is by no means entirely dependable in selecting for good teaching, it is probably the most useful criterion we have. Devoted scholarship would also be a criterion for good teaching (Morris Kline has much to say about scholarship, and we will be taking up that question later in this response), but we do not know how to produce good scholars in significant numbers except by precisely those stimuli which very successfully produce good research mathematicians.

The statement I have made above about the relation of successful research to good teaching is, I maintain, almost certainly true. On the other hand, being a very moderate claim, it is an unexciting statement. In this way it contrasts very markedly with the following statement made by Morris Kline: "There is in mathematics, and almost as surely in other disciplines, a direct conflict between teaching and research." Thus, Morris Kline has a great advantage in this dispute. It is the advantage which the extremist always has over the liberal. For the extremist is not inhibited in making extravagant, unqualified and dogmatic assertions which naturally catch the eye and the imagination of the audience. The liberal, more cautious in his search for truth, finds himself compelled to make only qualified claims – and nobody is going to man the barricades on the basis of a call to arms which explicitly recognizes some justice and some merit in the opposing cause!

Morris Kline used the same strategy very effectively in his earlier publication. He drew attention, with notable perspicacity, to some serious defects and unfortunate consequences of much of current mathematical education at the pre-college level. He then made a scathing attack on the community of research mathematicians, holding them almost uniquely responsible for everything that was unsatisfactory in the teaching of mathematics at the elementary and high school level. With resounding phrases

such as "professional mathematicians are the most serious threat to the life of mathematics, at least so far as the teaching of the subject is concerned" and "one of the basic reasons that mathematicians fail as pedagogues stems from the nature of the mathematical mind", he invited the public to join him in a campaign against the pernicious influence of his professional colleagues in a vital area of social concern. It was no wonder that his book was so successful, nor that, very exceptionally for a book devoted to the problems of mathematical education, it was reviewed – and favorably – by such influential journals as the New York Times and the Wall Street Journal. Morris Kline found himself quoted, doubtless to his own embarassment, by many who were opposed to the liberal policy of the federal government in support of scientific research. Those who wish to rebut Kline's view can expect to attract no such attention. The declaration that a charge of criminal and gross neglect of one's responsibilities is unfair is hardly newsworthy. Nevertheless, concern for our professional integrity as a community compels me to set down my reactions to Kline's fulminations and the basis for my assertion that they are, indeed, grossly unfair. It is for this reason that I agreed to respond to Kline's article in the Intelligencer, and I hope that my motives will be properly understood both by Kline and by the reader.[2]

Kline's Profile of the Research Mathematician

As we have said, Kline's thesis is that there exists a direct dichotomy between successful research and good teaching. Consistent with this thesis, he paints a picture in very sharp colors of the research mathematician and his attitude towards, and incompetence in, undergraduate teaching. Kline's profile of the research mathematician emerges very clearly from the following quotations from his book.

"For most of the teaching that the universities are, or should be, offering, the research professor is useless."

"What the professor does in his research has little if any bearing on what he has to teach at the undergraduate and even beginning graduate levels.... The creative researcher is most likely to be no more than a proficient but limited technician in one miniscule area."[3]

".... the enthusiastic professor who is deeply engrossed in his research.... rushes his students

toward it [i.e., *his speciality*] at a pace they cannot maintain and at the expense of far more valuable and basic material. He seeks to train specialists even though this training is useless to almost all of his students."

"Research requires an all-out effort and forces neglect, or perfunctoriness in teaching."

".... the research mathematician cannot devote time to such matters [i.e., *knowing what material is important for students and fashioning suitable courses*] and so teaches the material with which he is at ease."

"They [i.e., *researchers*] prepare 'perfect' lectures and regard interruptions as presumptuous or indicative of stupidity."

"Most professors are interested in mathematics and so cannot appreciate that there is a need to motivate students."

"Supplying effective motivation.... requires breadth of knowledge that the research-oriented professor does not have and may be unwilling to acquire."

"Scholarship is not easy or shallow as researchers would have us believe."

"The research professors are not at all interested in undergraduate activity."

There can be no doubt that Morris Kline can count on attracting attention by making such vast generalizations, such dogmatic assertions. In response, we can ask for evidence. We can also develop the argument to which I gave attention in the preceding section, explaining one view of the true nature of the positive relationship between successful research and good teaching. We must admit that there are outstanding mathematicians whose teaching is neglectful or perfunctory. But just as the presence of certain professional troublemakers on our campuses during the difficult days of student unrest did

---

[1] It was with the greatest reluctance that I abandoned my original intention to entitle this response *Why Morris Can't Learn.*

[2] In writing this response I only had available to me the excerpts published in the Mathematical Intelligencer, Vol. 1, 1 (1978), 5–14.

[3] Kline here attempts a generalization of his thesis adding, "A professor of English who has specialized in philology or in the love life of Madame Bovary is not by virtue of this research equipped to teach elementary composition or a survey of literature." The speciousness of the argument is evident. (a) It is not our contention that the *content* of research qualifies one for undergraduate teaching – it is the *impulse* to research activity which is likely to be conducive to inspiring teaching; (b) it is a caricature of research activity to describe a professor as specializing in the love life of a single character in a single novel. One might add that it is singularly odd for a professor of English to devote himself to the study of a French masterpiece!

not justify one in making sweeping and derogatory generalizations about student behavior and student attitudes, so the presence on our campuses today of some professors (by no means confined to the ranks of those active in research), who are unsuccessful teachers either through incompetence or neglect or both, does not begin to justify Kline's grotesquely sweeping attack.

Actually, the 'dilemma' which comes through most clearly from a perusal of the published excerpts of Kline's latest essay is not that to which Kline would have us pay attention, but *his* dilemma as to whether he should attack the research community for being bad mathematicians or for being good mathematicians! Some of the quotations already given suggest that Kline would be very happy to have the reader believe that practically all research mathematicians are bad mathematicians, at best narrow specialists grubbing around in the foothills of some particular and very esoteric domain of pure abstract mathematics. This point of view showed up clearly in Kline's earlier tirade, where he was very much concerned with those present-day mathematicians who denied the principal source of mathematical inspiration and who were thereby led into gross error in their recommendations for mathematics education. For Kline wrote, rather surprisingly for so respected an expert in intellectual history as he, "All the applications of mathematics to science come from mathematical ideas which were inspired by science." "Today's mathematicians", wrote Kline, "are no longer broadly educated in mathematics, to say nothing of science. . . . Most have turned to purely mathematical problems and to the formalization, axiomatization, and generalization of what is already known. Such tasks are far easier."

Slurs on the research mathematician of today abound even more in this latest work of Kline, where they are perhaps even more scurrilous. He writes, "The intensity of the competition even obliges researchers to adopt invidious practices more commonly attributed to the business world, such as refraining from telling others what they are trying to do or what partial success they may already have. For the sake of an immediate publication, a researcher may feel compelled to publish shoddy results, even though he may be quite sure that further efforts will produce a far better paper." It is surely not pedantic to say that by the use of the word 'obliges' in this context, Kline is deliberately creating the impression that the practices which he condems are common and *inevitable* among mathematicians today. Indeed, the imperative nature of the forces at work to corrupt the research mathematician is again expressed unambiguously by Kline in the following quotation: "There is but one human response to such a conflict. It is publish, and perish the students." Thus Kline charges the research mathematician with lack of scholarship, narrowness, dishonesty, and indifference to the needs of students!

Kline turns what is healthy in the approach of the research mathematician into something disgraceful. Thus "A professor's greatest joy is the bright doctoral student — and professors love to boast about their bright students

as though they made them bright." The disdainful attitude of researchers is expressed in Kline's characteristically dogmatic style. "But researchers do not deign to concern themselves with undergraduate or even graduate affairs that involve cooperation with other departments and the administration." And, finally, in this incomplete catalogue of calumnies on the mathematical community, we find the universities themselves pilloried precisely because, as previous quotations showed, it is, in Kline's view, the universities who are responsible for the unhealthy dominance of the research mathematician. Thus "One is quite safe in assuming that the more prestigious the university, the fewer its educational concerns and the poorer its educational effectiveness."

These quotations would lead one to suppose that Kline is at least single-minded in regarding present-day mathematics research as ill-conceived. This, however, turns out to be far from being the case. Perhaps with an eye on a different audience, or with a view to placating certain of his colleagues whom he would not wish to include in so scathing and comprehensive a condemnation, Kline devotes some attention to how we should make the best use of our research mathematicians! He shows an unexpected determination to make the life of a research mathematician as pleasant as possible, while, of course, ensuring that he does not have the opportunity to corrupt students! Thus, "Researchers who cannot or will not teach should be employed in institutes that are devoted entirely to research, should do no formal teaching and should give no degrees. In such institutions the researchers would be free to pursue their interests full-time. . . . The benefits to be derived from segregating excellent researchers who cannot or will not teach are immense. To burden such people with teaching is to deprive them of time and energy and distract them from their valuable work." Thus Kline is here recognizing the existence, and the right to existence, of research mathematicians, but he recommends that the vast majority of them should be isolated in research institutes, presumably without even graduate students, so that undergraduate students at our universities should be insulated against the disastrous effects of contact with them. It is difficult to understand why this country should accept the substantial cost of setting up these institutes if Kline's description of the characteristic activity of research mathematicians is indeed an accurate one. How many research mathematicians would in fact escape Kline's charge of narrowness and worthlessness? If what Kline has said about "the creative researcher" is fair, why should he be pampered and supported? The fact is that Kline is ambivalent and the reason is clear. Kline would still wish to encourage mathematical research that derives directly from problems in the physical sciences. To Kline this, and this alone, is the healthy part of mathematical research activity; the rest is corruption.

Nothing could more perfectly epitomize Kline's ambivalence than his final peroration. Having stigmatized the research mathematician in terms which at least have

the merit of being totally unequivocal, having recommended the segregation of research mathematicians in institutes entirely away from students, Kline finishes with the clarion call, "Only recognition of the interdependence of research, scholarship and teaching can advance mathematics itself, improve teaching, and further the multitudinous valuable uses of mathematics in our society." Humpty-Dumpty rides again!

From the quotations I have given, Kline's strategy to convince his reader emerges very clearly. There are legitimate dissatisfactions with the quality of much of the instruction which students receive today. Of course there has always been such legitimate dissatisfaction, but I dare to assert that there has never been a more conscientious attempt to improve the quality of education than one finds today. Kline, employing a very impressive mastery of polemic, seeks to convince the reader that he has identified the villain of the piece. He does this by hammering away with dogmatic assertions and massive generalizations. It would not be possible to argue categorically that research and teaching were incompatible — and, to Kline's credit, he is perfectly explicit that this is his thesis — if there were research mathematicians who were good teachers. Thus, Kline is forced into denying this possibility. He reinforces his technique of argument by quotations of the views of authorities which he then takes to be statements of empirical fact. One quotation on our part should suffice to demonstrate the nature of this method.

Kline quotes Bertrand Russell's view of the motivation and attitudes of mathematicians, "Remote from human passions, remote even from the pitiful facts of nature. . . . ", and refers to Bertrand Russell's statement that he was drawn into mathematics "because it is not human". Bertrand Russell was a great philosopher and, in my view, a most remarkable and wonderful man; but he was not a creative mathematician. He was, as a logician of those days, a critic of mathematics. Are we then to accept his view about mathematical activity as being beyond question? Were Archimedes, Newton and Gauss not mathematicians? Kline's method is simply to accept Bertrand Russell's view as being holy writ, expressing unarguable truths, and so he follows his quotations from Russell with his own categorical statement, "Mathematics then is a refuge. Research. . . . is likely to be a haven for those with scanty spirit and interests." This might follow if one accepted Russell's premises. As an empirical truth, it is laughable, contradicted by the experience of everybody, Kline included, who has had any serious contact with the mathematical community. Thus, here as elsewhere, an argument having an immediate, superficial appeal, is found on closer analysis to be based on hidden assumptions of the infallibility of carefully chosen authorities; we do not, on closer inspection of the argument, accept the assumptions and have consequently no need to accept the ridiculous conclusions.

Kline's View of the Scholar

It would appear from the excerpts available to me in writing this reply that Kline's solution to the problems which he sees confronting undergraduate mathematics education would be the emergence of a new type of mathematician, the scholar. The characteristic qualities of the scholar appear to be an ability to assess the significance of mathematical research and to relate that research to activity in allied disciplines and to mathematical education. Kline claims that we have practically no mathematical scholars. "Because scholarship, as opposed to research, is not valued, professors do not cultivate it. There is no place for the scholar in the mathematical world." The qualities which Kline asks of the scholar are formidable indeed. It is very rarely that one is able to assess, immediately or even within a short space of time, the relevance and importance of a substantial piece of mathematical research, and few indeed would ever be able to do this. It is true that one can sometimes say that a piece of research is useless rather quickly, but it is much, much rarer to be able to say, without the lapse of much time and the study of numerous sequels to that research, that a piece of research is of great value. Moreover, the people who can say this with any authority are almost exclusively those who are themselves active in the field in which the research lies, or in a closely-related field. This is not because there are no scholars in Kline's sense — it is because the process of evaluating progress in mathematics is a natural part of the process of research itself. If we knew how to produce scholars of the kind that Kline asks for, then I am sure we would do so, and we would be happy to have them as colleagues. The only outstanding scholars known to me are all of them simultaneously active in research. It is no accident that almost all of them are outstanding teachers.

There is one respect in which Kline's recommendation for the encouragement of the scholar cannot command our respect. For Kline, taking every possible opportunity to bring current mathematical research into disrepute, sees as one of the tasks of the scholar that of monitoring the dishonesty of his research colleagues. This is perhaps the most unpleasant slur of all that one finds in Kline's whole essay. What should we think of a mathematician who can write of the scholars' role ". . . . They also can detect duplication of results; and the very knowledge that such duplication will be detected will deter those who *would consciously publish old results disguised in new terminology and symbolism".* What a detestable arrangement Kline appears to be envisaging — and how utterly unnecessary in view of the extreme rarity of deliberate plagiarism among mathematicians. Let us be magnanimous and assume that Kline would not be held to the implications of what he there wrote — all that we know from personal experience suggests that he is far too honorable to accept those implications himself. Let us rather return to the better aspects of Kline's argument for the scholar.

Our experience makes it difficult to avoid the conclusion that Kline's recommendation that we should encourage the emergence of scholars who assess research rather than do research is entirely unrealistic. Let us agree with him that we should value scholarship – indeed my only disagreement with him is in his statement that we do not. But let us also point out that a recommendation that would require us to substantially increase the supply of people who were simultaneously throughly familiar with research and conscientious and successful teachers comes ill from one who has committed himself explicitly, dogmatically, and unequivocally to the view that there is an ineluctable conflict between mathematical research and education.

Is There a Problem? Is There a Solution?

Just as with respect to Kline's earlier book, he could have done a real service to the community of researchers, scholars and teachers of mathematics had he been content to draw attention to the existence of serious problems. Unfortunately, it is difficult to avoid the conclusion that this was only a very small part of Kline's purpose. Plainly, his main purpose is to lambast the research establishment – that is where his heart lies! Thus if we are to salvage something of value from Kline's immoderate attack, we must first ask ourselves the question with regard to undergraduate education and the structure of our universities – Is there a problem? And we must then ask ourselves the consequent question – Is there a solution?

In speaking of a problem, I do not refer simply to the problems which are the continuing concern of undergraduate teachers of mathematics – rendering the teaching more effective, improving the curriculum, enriching the options, and so forth. I, myself, am currently very much involved in these 'problems', having in a moment of weakness accepted the invitation to chair the NAS-NRC Committee on Applied Mathematics Training. I have learnt a great deal from listening to the deliberations of my colleagues on this committee, and believe that we can contribute to improvement in the structure and content of, for example, the undergraduate mathematics major. Difficult as such questions are, they are not so intractable as to be labeled problems in the sense in which Kline has identified a problem.

I believe it can be agreed that there is a problem in this severer sense. It is the problem of the appropriate recognition of good teaching in an institution of higher education and learning. Particularly today, when there are so very few jobs to be had in the academic world, the criteria for selecting faculty members, for promoting them, for giving them tenure, and for giving them salary raises become very crucial and delicate matters. Plainly these criteria depend to some extent on the nature of the institution. That is to say, it would surely be common ground that, in a four-year college, there should not be so much emphasis on research as in a university boasting a good graduate school. There is the problem of what to do about the bad teacher, especially when he has tenure, and this problem is particularly acute when he is an excellent research mathematician. This problem is difficult, but I do not believe it is a characteristic problem facing us today. Much more typical is the problem of deciding how to esteem and reward the faculty member who performs creditably if not outstandingly in all aspects of his professional responsibilities. The problem is aggravated by the fact that, although we have good objective criteria for deciding the quality of a person's research – I refer to peer evaluation – we do not have the same reliable means of assessing teaching skill and effectiveness.[4] Thus the problem of rewarding good teaching is compounded by the problem of identifying it.

Earlier in this response I indicated certain criteria of good teaching, and I do believe that these criteria could indeed be applied in order to identify the good teacher. Let me therefore leave the problem of identification and assume, for the sake of argument, that it can be solved. I would then like to suggest that it is proper to reward good teaching and that the appropriate reward should be salary increase rather than promotion. I believe that promotion and tenure should be the reward of outstanding work of an imaginative and innovative nature. Such outstanding work could be in the field of mathematical research, but does not have to be. Thus it is perfectly possible – and, today, more important than ever – to show imagination, energy and enterprise in the development of new courses and the modernization of old ones. If such outstanding work in the field of education were rewarded in a way comparable with outstanding research, then I believe we would have found a solution to what is undoubtedly a serious problem. If Kline had used his great gifts as mathematician and expositor to open a serious debate on this problem, we would all have been in his debt. But he has chosen instead to close the debate by proposing to take research out of the universities. Such a proposal is entirely unacceptable; we must, on the contrary, ensure that good research contines to be fostered and encouraged alongside good teaching in our institutions of higher learning. While suitably rewarding good teaching and outstanding enterprise with regard to curriculum and education, we must resist with all our strength any suggestion that good research should cease to be a criterion for promotion and tenure. Any university which adopts that suggestion has taken a giant stride along the road to mediocrity. And it is well to remember that the road to mediocrity is steep and downhill. Descent, once initiated, will be rapid and accelerating; and there is no guarantee that one can find the means to climb back up.

[4]There is a mysterious paragraph on page 8 of Kline's excerpted article in which he appears to be arguing that department chairmen would be more reliable than outside specialists in judging the research potential of a young mathematician. However, Kline also says that many chairmen do not possess this ability – and at the same time stigmatizes the outside specialist "who may be biased". Precisely who comes off well in this part of Kline's analysis is not clear! I do not share Kline's prejudice against the outside specialist; and I do not think it fair to demand so much of departmental chairmen. I regard the refereeing system as broadly effective.

Battelle Memorial Institute
Seattle, Washington, USA

Riemann's Example of A Continuous "Nondifferentiable" Function Continued

S. L. Segal

Under the first seven words of the above title Doctor Erwin Neuenschwander recently published an interesting and informative article in the Intelligencer [8] which, unfortunately, is not completely up-to-date in some of its statements.

The function in question is

$$\sum_{n=1}^{\infty} \frac{\sin n^2 x}{n^2}$$

According to Weierstrass (in a letter of 1873 to Paul duBois-Reymond) [10] " . . . Riemann already in the year 1861 characterized the function represented by the series

$$\sum_{n=1}^{\infty} \frac{\sin n^2 x}{n^2}$$

to some of the attenders at his lectures as one which had no derivative; however, his proof was told to no one; on the contrary, he only indicated in passing that it was to be obtained from the theory of elliptic functions."

Doctor Neuenschwander gives a fascinating discussion of this "Riemann Function" and related concerns, including material from the "Casorati Nachlass" recently discovered by him. As he also mentions, Hardy [6] was apparently the first to take up Riemann's purported statement with any success. leaving some cases open which were settled by Gerver. Unfortunately, and somewhat curiously, he stops short of realizing that, in fact, Gerver [3], [4], [5] managed to settle *all* cases left open by Hardy. An extremely simple proof of the Hardy-Gerver results depending on a number-theoretic argument has since been given by A. Smith [9], who actually shows slightly more.

Smith's paper was motivated by the wish to settle the cases left open by Gerver in [4] as well as to provide a proof simpler than that lengthy paper. The only techniques used by Smith are the Poisson summation formula and results on Gaussian sums. Thus there is now a reasonably simple proof that this "Riemann function" has no finite derivative at any point other than those of the form x = π (2A + 1) / (2B + 1) (A, B integers)

where it has derivative – ½. That this proof (or Gerver's for that matter) would seem well within Riemann's ken has no bearing on his purported statement concerning non-differentiability. In any case, no proof utilizing elliptic functions (or even indicating their relevance) is known.

In fact, Weierstrass, in the cited letter to DuBois-Reymond [10], doubts whether Riemann thought his function failed to have a finite derivative at any point, for, as he says, " . . . in the circle of Riemann's pupils, at least, such functions seem not to have been known", citing remarks of Hankel in support. On the other hand, in the published paper [11], Weierstrass gives his opinion that such functions *were* what Riemann was considering. It is perhaps also worth noting that, in a letter of 1876 to Koenigsberger [12], Weierstrass states that his function

$$\sum_{n=0}^{\infty} b^n \cos(a^n \pi x), \quad a > 1, \quad \text{and} \quad \frac{1}{a} < b < 1$$

is nowhere differentiable (a result usually attributed to Hardy [6] as Weierstrass only published the result for a an odd integer and ab > 1 + 3π / 2 [11]), and discusses when the result of nondifferentiability holds for ab = 1. In the same letter, Weierstrass goes on to conjecture about conditions on Fourier cosine series which produce continuous nondifferentiable functions.

A few concluding remarks:

(1) Gerver was apparently a student at Columbia who was stimulated to consider the "Riemann function" by a paper of Kahane [7] mentioning the open problems connected with it. He seems to have done no published mathematics other than the ensuing announcement and two papers, [3], [4], [5].*

(2) K.M. Garg has remarked [2] that the fact that the Riemann function has derivative + ∞ at 0 has long been known.

(3) Define {t} by the well-known Fourier series

$$\frac{-1}{\pi} \sum_{m=1}^{\infty} \frac{\sin 2m \pi t}{m} = \{t\}$$

Let λ (n) = (-1)$^{\Omega (n)}$ where Ω (n) is the total number of prime factors of n (multiple factors counted multiply). Then Davenport [1] proved that

$$-\pi \sum_{n=1}^{\infty} \frac{\lambda(n)}{n} \left\{\frac{nx}{2\pi}\right\} = \sum_{n=1}^{\infty} \frac{\sin n^2 x}{n^2}$$

Added in proof: The Intelligencer recently received a letter from Gerver, who is now at the University of Hawaii.

where the convergence of the series on the left is uniform in x. While this formula (incidental to Davenport's main purpose) is fascinating, it is hard to see how it might be used in studying the differential behavior of the function on the right.

REFERENCES

[1] Davenport, H., On some infinite series involving arithmetic functions, Quarterly Journal of Mathematics 8 (1937), pp. 8–13 and part II, same volume, pp. 313–320.

[2] Garg, K. M., Mathematical Reviews, vol. 46 (1937) Review No.7451 (p. 1286, Review of [9]).

[3] Gerver, J., The Differentiability of the Riemann Fuction at Certain Rational Multiples of π, Proc. Nat. Acad. Sci. 62, (1969), pp. 668–670.

[4] Gerver, J., The Differentiability of the Riemann Function at Certain Rational Multiples of π, Amer. J. Math 92, (1970), pp. 33–55.

[5] Gerver, J., More on the Differentiability of the Riemann Funktion at Certain Rational Multiples of π, Amer. J. of Math. 93 (1971), pp. 33–41.

[6] Hardy, G. H., Weierstrass's Non-Differentiable Function, Transactions Am. Math. Soc. 17, (1916), pp. 301–325.

[7] Kahane, J. P., Lacunary Taylor and Fourier Series, Bull. Am. Math. Soc. 70, (1964), pp. 199–213.

[8] Neuenschwander, E., Riemann's Example of a Continuous, "Nondifferentiable" Function, The Mathematical Intelligencer 1, (1978), pp. 40–44.

[9] Smith, A., The Differentiability of Riemann's Function, Proc. Am. Math. Soc. 34 (1972), pp. 463–468.

[10] Weierstrass, K., Letter of 23 November 1873 to Paul DuBois-Reymond, reprinted in Acta Mathematica, 39, (1923), pp. 199–201.

[11] Weierstrass, K., über continuirliche Functionen eines Reellen Arguments, die für keinen Werth des letzteren einen bestimmten Differentialquotienten besitzen, K. Weierstrass, Mathematische Werke II, pp. 71–74 (paper read in the Academy of Sciences 18 July (1872).

[12] Weierstrass, K., Letter of 10 February 1876 to L. Koenigsberger reprinted in Acta Mathematica 39 (1923), pp. 231–234.

University of Rochester
Rochester, New York
USA

****

Obituary

I. N. Vekua (1907-1977)

The Soviet mathematician Ilya Nestorovich Vekua died recently at the age of 70, after a long illness. Vekua's mathematical work was mainly devoted to exploring the interplay between complex analysis and elliptic partial differential equations and to applying this interplay to the theory of elasticity. In his research on elasticity, Vekua continued the tradition established by his teacher and friend N.I. Muskhelishvili.

In the theory of elliptic partial differential equations in the plane Vekua followed two different paths. One, similar to the work of Stefan Bergman, led to explicit formulas for solutions of elliptic equations with real analytic coefficients, in terms of arbitrary holomorphic functions of a complex variable. Vekua applied his formulas to very general boundary value problems and, in this connection, obtained a result which is today recognizable as a precursor of the Atiyah-Singer index theorem.

The other main direction of Vekua's investigations concerned itself with properties of systems of two first order elliptic equations, viewed as generalized Cauchy-Riemann equations. This time real analyticity of the coefficients is not assumed, and it turns out that the properties in question are best understood in terms of a generalized complex function theory. While Vekua's "generalized analytic functions," are basically the same as my "pseudoanalytic functions," our motivation and our aims were somewhat different. In particular, Vekua applied his theory to elastic shells and to infinitesimal deformations of surfaces in 3-space.

Of course, he worked also on many other questions. For instance, he was the first to apply the Calderon-Zygmund inequality for singular integrals to solving Beltrami's equations. A detailed report about Vekua's research, and a complete bibliography, will be found in *Uspekhi 12* (1957), pp. 227-234, *22* (1967), pp. 185-195 and *32* (1977), pp. 2-21.

Vekua was an influential member of the Soviet scientific establishment and was given many responsibilities and honors. He was a member of the USSR Academy of Sciences and of its Presidium, and, at his death, President of the Georgian Academy. He was the first President of the University of Novosibirsk and later headed the University of Tbilisi. He was also elected to the Supreme Soviet of the USSR and to the Central Committee of the Communist Party of Georgia.

But there was nothing official or rigid in his personality. Ilya Nestorovich was a large man with a booming voice, enormously friendly and filled with *joie de vivre*. His courtliness was somewhat old fashioned, and when you saw him kissing ladies' hands, or officiating as "tamada" (toastmaster) at a Georgian style party, or enjoying food and good wine or a good story, you forgot all about his scientific and administrative prominence. But his benevolence went beyond mere social friendliness. He was willing to help when others would not, and one could talk to him, freely and openly, about many things. His judgements were shrewd, original, and unhampered.

We met first in 1958 at the International Congress of Mathematicians in Edinburgh. For the first time a sizeable group of Soviet mathematicians came to a scientific gathering in the West, and neither the Americans nor the Soviets quite knew how to behave to each other. The ice was broken by Vekua — he invited everyone to a big party in his hotel room.

Vekua's deeply rooted optimism was not based on denying or ignoring the tragic aspects of life or of history. Once he and I watched a Broadway performance of Bolt's *Man For All Seasons*. Vekua may not have known the details of Sir Thomas More's death, and the dilemma of an honorable and reasonable man confronting a tyrant moved him deeply. I recalled that evening when I read in an article, written for his 70th birthday *(Uspekhi, 32)* about the "strength, courage and self-control Ilya Nestorovich showed under complicated circumstances."

Lipman Bers
Columbia University
New York, USA

*****

Maximum Antichains in the Partition Lattice

R. L. Graham

Many questions in combinatorics deal with collections of subsets of a finite set. One of the most basic results of this type, first proved by E. Sperner [16] in 1928, asserts the following: If $S_1, S_2, \ldots, S_t$ are subsets of an n-element set S such that no S_i is a subset of any other S_j then

$$t \leq \binom{n}{\lceil \frac{n}{2} \rceil}$$

Furthermore, this bound on t can only be achieved by taking the S_k to be all the $\lceil \frac{n}{2} \rceil$-element subsets of S, or, when n is odd, by taking all the $\left(\lceil \frac{n}{2} \rceil + 1\right)$-element subsets.

Sperner's theorem, like Schur's work [15] on the solutions of $x^m + y^m = z^m$ van der Waerden's theorem [17] on arithmetic progressions, Ramsey's fundamental result [12] on partitions of the subsets of a set, and Polya's approach [11] to the theory of enumeration, has been the seed from which a major branch of combinatorial theory has grown during the past 50 years. This branch, often called extremal set theory, has been especially active during the past 10 years. In particular, one of the outstanding open problems, which was first raised by G.–C. Rota nearly 15 years ago and which was responsible for much of this activity, has just been settled within the past year by E. Rodney Canfield of the University of Georgia. What is even more intriguing is that Canfield showed that the answer that everyone had expected (and was trying to prove) was wrong*.

In this note I would like to give a brief sketch of the background of Rota's problem and its resolution.

By a chain C in a (finite) partially ordered set P we mean a totally ordered subset of P; the length of C is just the number of elements in it. We say that P is *graded* if P has a unique minimal element 0 and for every $p \in P$, all maximal chains from 0 to p have the same length, called the *rank* of p. We denote the elements of P having rank k by R_k, also called the kth level. By an *antichain* in P we mean a subset of mutually incomparable elements of P. For example, for any k the set R_k forms an antichain.

*This often explains the difficulty encountered in trying to prove a result.

What Sperner's theorem says is that for $P = S_n$, the class of subsets of an n-element set S partially ordered by inclusion, the largest antichain must be of this form for the choice of k which maximizes $|R_k|$ (in this case $k = \lceil \frac{n}{2} \rceil$).

It is natural to ask to what extent this result holds for more general graded partially ordered sets. In particular, G.–C. Rota [14] raised the question of whether this is true for P_n, the lattice of partitions of an n-element set. The elements of P_n of rank k, denoted by R(n,k), are the partitions of a fixed n-element set, say $\{1, 2, \ldots, n\}$, into k nonempty blocks. We say that the partition π *refines* the partition π' if each block of π is contained in a block of π'. P_n is partially ordered by refinement. In Fig. 1 we show P_4.
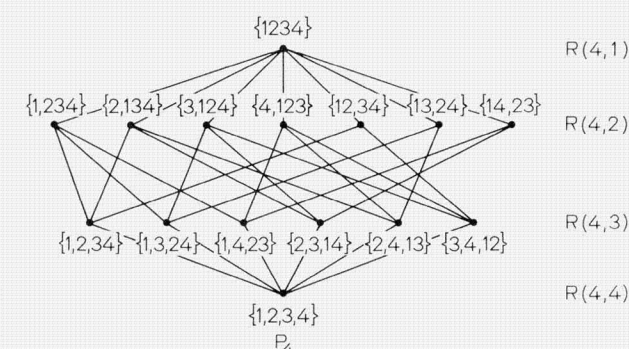
Figure 1

The number of elements of rank k is denoted by S(n, k). These numbers are known as the Stirling numbers of the second kind [13]. They satisfy the simple recurrence

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k)$$

and although they have been known for hundreds of years, there are still some surprising gaps in our knowledge about them. For example, it is not even known whether S(n, k) = S(n, k+1) can ever hold for n > 2 (it can be shown that S(n, k) = S(n, k+1) = S(n, k+2) is impossible). It is known however that the S(n, k) are *unimodal*, i.e., S(n, k) — S(n, k+1) has only one change of sign as k goes from 1 to n — 1. It follows from this that Rota's question is equivalent to showing that between any two consecutive levels R(n, k) and R(n, k+1) of P_n, one can always find what is called a *matching*. This is simply a one-to-one pairing of each partition π of the smaller of the two levels with a distinct *comparable* partition π' of the larger of the two levels. For example, one such matching between R(4,2) and R(4,3) in Fig. 1 is given by:

$$\{1,234\} \;-\; \{1,3,24\} \;,\; \{2,134\} \;-\; \{1,2,34\} \;,$$
$$\{3,124\} \;-\; \{2,3,14\} \;,\; \{4,123\} \;-\; \{1,4,23\} \;,$$
$$\{12,34\} \;-\; \{3,4,12\} \;,\; \{13,24\} \;-\; \{2,4,13\} \;.$$

It is easy to see that the existence of these matchings implies that the maximum antichains of P_n have $\max_k S(n,k)$ elements. For if we have an alleged maximum antichain A having more than $\max_k S(n,k)$ elements, say A has elements below the largest level, simply replace each of A's lowest rank elements π by the (comparable) partition π' paired with it in the adjacent level above (a similar argument applies if A has elements above a maximum level). The new set A' is still an antichain, has as many elements as A and has its elements in one less level. Using the unimodality of the S(n, k), we can by induction find an antichain belonging entirely to a largest level, which is a contradiction.

Thus, Rota's problem is reduced to deciding whether there is always a matching between R(n,k) and R(n, k+1). The most well-known tool for proving the existence of matchings is the so-called Marriage Theorem of P. Hall [8]. This result applies to the following situation. Suppose A and B are sets of men and women, respectively, and each man $a \in A$ is acquainted with a set of women $T(a) \subseteq B$. We can conveniently represent this by a bipartite graph where an edge between a and b indicates acquaintanceship (and we assume without loss of generality that $|A| \leq |B|$). Then a necessary and sufficient condition that each man be able to marry a women he knows is that for all k, each set of k men know altogether at least k women. In other words, there is a matching between A and B iff for all $X \subseteq A$,

$$\left| \bigcup_{x \in X} T(x) \right| \geq |X| \quad .$$

Unfortunately, unless there is a fair amount of regularity in the bipartite graphs in question, the use of Hall's theorem can be impractical in proving the existence of matchings (since it requires verifying $2^{|A|}$ conditions).

Other approaches to Rota's problem during the past 10 years resulted in the introduction of a variety of fruitful concepts and significant results into matching theory, such as the idea of a normalized matching [5], Harper's product theorem [9], and the far-reaching work of Greene and Kleitman [6], [7] and Katona [10], but the original problem itself still stood relatively untouched. It was shown that any antichain in P_n had at most $\max_k S(n, k)$ elements for $n \leq 20$ (and even this was not completely trivial because of the size of the S(n, k)'s, e.g., S(20,8) = 15170932662679) and it was generally believed this would continue to hold for all n.

Thus, it was quite unexpected when Canfield showed that there are antichains in P_n having many *more* than $\max_k S(n, k)$ elements when n becomes very large.

Basically, his technique for proving this involved:
(i) determining to within one the value k_n for which $S(n, k_n) = \max_k S(n, k)$;
(ii) defining a special class C of partitions in the level $R(n,k_n-1)$; more specifically, C consists of all $\pi \in R(n,k_n-1)$ having exactly t blocks of size $\leq m$ and the remaining $k_n - t - 1$ blocks of size $> m$ and $\leq 2m$, where t and m are appropriately chosen;
(iii) obtaining precise estimates of the size of C and the size of span (C), defined to be the set of all π' $\in R(n, k_n)$ which are refinements of some $\pi \in C$; this is done by using local limit theorems and the Berry-Esséen inequality for the estimation of coefficients of certain polynomials;
(iv) using these estimates to show that $|Span(C)| < |C|$.
Thus

$$R(n,k_n) - Span (C) \cup C$$

is an antichain in P_n with more than $S(n, k_n)$ elements (see Fig. 2). The details can be found in [1], [2], [3], [4].
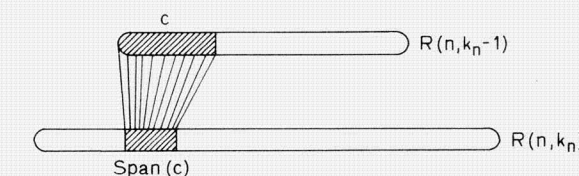
Figure 2

One might well ask just where these large antichains are located in P_n and, in particular, what the least n is for which P_n has antichains with more than $S(n, k_n)$ elements.

As to the first question, Canfield has shown that the maximum antichains A_n in P_n for large n must contain elements in levels quite far from the largest level $R(n,k_n)$. In fact, for any $\epsilon > 0$, A_n must have elements in both $R(n, k_n+t_1)$ and $R(n,k_n-t_2)$ for some $t_1, t_2 > \sqrt{n}/(\log n)^{1+\epsilon}$ when n is sufficiently large.

As to the second question, Canfield estimates that his techniques will start working at about n = 6.526x10²⁴. At this value of n, $S(n, k_n)$ can be enormous, e.g., exceeding

$$10^{10^{20}} .$$

Thus, it is conceivable that we will never know the first time P_n has an antichain with more than $\max_k S(n, k)$ elements!

REFERENCES

[1] E. Rodney Canfield, On a problem of Rota, Bull. Amer. Math. Soc., 84 (1978), p. 164.

[2] ————, Application of the Berry-Esséen inequality to combinatorial estimates, (1977) (preprint).

[3] ————, On the location of the maximum Stirling number(s) of the second kind, (1977) (preprint).

[4] ————, On a problem of Rota, (1977) (preprint).

[5] R. L. Graham and L. H. Harper, Some results on matching in bipartite graphs, SIAM Jour. on Appl. Math., 17 (1969), 1017–1022.

[6] C. Greene, Sperner families and partitions of a partially ordered set, Combinatorics, Math. Centre Tract 56, (1974) Amsterdam, 91–106.

[7] C. Greene and D. J. Kleitman, On the structure of Sperner k-families, Jour. Comb. Th. (A), 20 (1976), 41–68.

[8] P. Hall, On representatives of subsets, Jour. London Math. Soc., 10 (1935), 26–30.

[9] L. H. Harper, The morphology of partially ordered sets, Jour. Comb. Th. (A), 17 (1974), 44–58.

[10] G. O. H. Katona, Extremal problems for hypergraphs, Combinatorics, Math. Centre Tract 56, (1974) Amsterdam 13–42.

[11] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, Acta Math. 68 (1937) 145–254.

[12] F. P. Ramsey, On a problem of formal logic, Proc. London Math. Soc., 2nd ser., 30 (1930) 264–286.

[13] J. Riordan, An introduction to combinatorial analysis, Wiley, New York, 1958.

[14] G. –C. Rota, Research problem 2-1, Jour. Comb. Th. 2 (1967), p. 104.

[15] I. Schur, Über die Kongruenz $x^m + y^m = z^m$ (mod p) Jahrb. Deutsch Math. – Verein. 25 (1916) 114 – 116

[16] E. S. Sperner, Ein Satz über Untermengen einer endlichen Menge, Math. Z. 27 (1928), 544–548.

[17] B. L. van der Waerden, Beweis einer Baudetschen Vermutung, Nieuw Archief voor Wiskunde, 15 (1927) 212–216.

Bell Laboratories
Murray Hill, New Jersey, USA

INTERNATIONAL CONGRESS OF MATHEMATICIANS

15-23 August 1978—Helsinki, Finland

One-hour plenary addresses

L.V. Ahlfors	Quasiconformal mappings, Teichmüller spaces, and Kleinian groups
A.P. Calderón	Commutators, singular integrals on Lipschitz curves and applications
A. Connes	On the classification of von Neumann algebras
R. Dobrushin	Classical statistical mechanics as a branch of probability theory
R.D. Edwards	The topology of manifolds and cell-like maps
D. Gorenstein	The classification of finite simple groups
M. Kashiwara	Micro-local analysis
N.N. Krasovskiĭ	Control under uncertain information and differential games
R. Langlands	Automorphic representations and L-functions
Y. Manin	Modular forms and number theory
S. Novikov	Linear operators and integrable Hamiltonian systems
R. Penrose	The complex geometry of the natural world
W. Schmid	Representations of semisimple Lie groups
A.N. Shiryayev	On absolute continuity and singularity of probability measures on functional spaces
W.P. Thurston	Geometry and topology in dimension three
A. Weil	History of mathematics: why and how
S.-T. Yau	The role of partial differential equations in differential geometry

---

News

Wolf Prizes Awarded

I.M. Gelfand of Moscow State University and C.L. Siegel of the University of Göttingen were jointly awarded the Wolf Prize in Mathematics April 10 at the Knesset in Jerusalem. The mathematics prize and the simultaneous awards in physics, chemistry, medicine, and agriculture were the first scientific Wolf Foundation prizes. Each prize bears a $100,000 grant, to be divided equally in the case of more than one recipient in a single field. The nominating committee in each field consists of three world-renowned scientists, with one member of each committee being Israeli.

The purposes of the Foundation are to promote science and the arts, to grant scholarships to students and loans to universities, and to award prizes to well-known scientists and artists for their achievements on behalf of humanity and on behalf of friendly relations among peoples irrespective of citizenship, race, color, religion, sex, or political outlook. The Foundation's Council is chaired by the Israeli Minister of Education and Culture. Its funds of $10 million have been donated anonymously.

Completing this year's prize list are: C.-S. Wu, Columbia University, for physics; C. Djessari, Stanford University, for chemistry; G.D. Snell, Jackson Laboratory, Bar Harbor Maine, J. Dausset, St. Louis Hospital, Paris, and J.J. van Rood, University of Leiden, for medicine; and G.F. Sprague, University of Illinois, and J.C. Walker, University of Wisconsin, for agriculture.

U.S. Mathematics Research Institute Approved

The National Science Board, whose members are scientists appointed by the President and which, together with the Director, constitutes the National Science Foundation, after extensive discussion passed the following resolution on March 16: *Resolved that the National Science Board approves the establishment of a Mathematical Research Institute; further, the Board approves of a Project Announcement and the general plans for the establishment of the Institute.* This means that a Project Announcement and a call for proposals for the design of the Institute will shortly be circulated widely – to all university presidents, offices of research, and interested industrial concerns. The nature and sources of funding for the Institute are open questions and include the possibility of non-governmental participation. Another major question now opened for discussion is location. For information on the considerations that were involved in formulating plans for the Institute-to-be, see the News in Number 1.

Development of a Common DoD Programming Language

The U.S. Defense Department has been engaged in a project to develop a common high-order programming language for its own use. The target is to produce a language appropriate for use in all "embedded system" applications. This includes applications like command and control, communications, real time control systems, system software such as compilers and operating systems, and so on. The currently envisaged programming activities exclude major business applications, which tend to use COBAL, and scientific applications, which more often use FORTRAN, though it seems possible that the resultant language may have some impact on those fields also. Embedded system programming is reputed to cost the Defense Department more than three billion dollars per year, much of it redundant or inefficient. The expectation is that common use of the new language will make this activity significantly more efficient.

DoD requirements were consolidated in a series of documents culminating in "Ironman (Revised)" of July 1977. This was taken as the specification for competitive prototype design contracts. Reports were received from four contractors on February 15, containing detailed descriptions of their preliminary designs. All four are using PASCAL as the basis of their work. DoD solicited analyses of these designs, received in mid March, from about 130 different organizations. They plan to continue some of the contracts to the final language specification to be completed by about this time next year when a single selection will be made and contracts will be let for software to implement the language. By some time in 1980 it is foreseen that the common language will be widely implemented and available to DoD programmers and programming contractors in the embedded systems area.

The DoD language commonality effort has wide-ranging implications and will certainly have an impact on the computer industry as a whole. It means that the system software industry, which already is showing strong signs of shifting to PASCAL, will be given a huge incentive to shift even further in order to be positioned to sell software and services to the Defense Department.

Survey Provokes Protest

Controversy has emerged over *The 1977 Survey of the American Professoriat*, so named by its directors, E.C. Ladd, Jr. (Social Science Data Center of the University of Connecticut at Storrs) and S.M. Lipset (Hoover Institute of Stanford University). The survey consists of a 128-item questionnaire mailed last spring to approximately 9000 faculty members at US colleges and universities. S. Lang (Department of Mathematics, Yale University) was a recipient, refused to answer the questionnaire, and encouraged by N. Koblitz and J. Tate, wrote a letter of protest to Lipset. He was soon joined by others, including a number of mathematicians and several sociologists, who protested the survey in letters to its directors and to various private and governmental agencies which have been thought to have connections with it.

These critics charge that Ladd and Lipset disclosed neither their sources of financial support nor how the survey results would be used to bring faculty opinion into the formulation of educational policy – this being the surveyors' stated aim. Beyond these procedural criticisms, they claim that the survey contains questions, for example, on US military policy and on sexual morality, that have no apparent connection with educational policy. In questions that do bear on educational policy, dealing with such phenomena as grade "inflation" and collective bargaining by faculty, the multiple-choice answers have been constructed, they feel, so as to prejudice the responses. Their overall impression of the questionnaire is that its choices have often been formulated as hypothetical answers to hypothetical problems designed more to draw a political characterization of the respondent on a one-dimensional scale from liberal to conservative than to solicit solutions to actual policy problems. Critics who view the survey as more or less overtly political have referred to another controversy engaging Lipset, this one conducted in the columns of *The New York Review of Books* last spring. There, S. Diamond, a Columbia University sociologist, wrote a highly critical review of Lipset's essay in the book *Education and Politics at Harvard*. Lipset's essay deals with the administrative policies of Harvard University during the McCarthy period; Diamond, whose review stimulated a long exchange in the letters column, described the effect of those policies in his own case.

Lipset's and Ladd's analyses of the survey results have been appearing since September in the *Chronicle of Higher Education*. Lang's critique was published May 18 in *The New York Review of Books*.

New and Noteworthy

Following is a selection of books that have recently been published, or, in a few cases, reissued. Suggestions for contributions to the list are heartily welcome and should be received by the editors two months prior to publication of the Intelligencer. Please note that titles of volumes of conference proceedings are here given with the subject area of the conference first, for ease of reference. In certain cases this permutes the words of the actual title which may begin, for example, with "proceedings".

Alexander, J.P., P.E. Conner, and G.C. Hamrick: Odd Order Group Actions and Witt Classification of Inner Products. Lecture Notes in Mathematics, Vol. 625. Springer-Verlag, 1977.

Arnold, D.M., R.H. Hunter, and E.A. Walker (Eds.): Abelian Group Theory. Proceedings of the New Mexico State University Conference, Las Cruces, 1976. Lecture Notes in Mathematics, Vol. 616. Springer-Verlag, 1977.

Aubin, J.: Applied Abstract Analysis. John Wiley, 1977.

Baily, W.L. and T. Shioda (Eds.): Complex Analysis And Algebraic Geometry: A Collection of Papers Dedicated to K. Kodaira. Cambridge University Press and Iwanami Shoten, 1978.

Baker, A. and D.W. Masser (Eds.): Transcendence Theory: Advances and Applications. Proceedings of the Cambridge Conference, 1976. Academic Press, 1977.

Baker, C.T.H.: The Numerical Treatment of Integral Equations. Oxford University Press, 1978.

Ball, J.A.: Factorization and Model Theory for Contraction Operators with Unitary Part. Memoirs, Vol. 198. American Mathematical Society, 1977.

Barut, A. and R. Raczka: Theory of Group Representations and Applications. Polish Scientific Publishers, 1977.

Baues, H.J.: Obstruction Theory: On Homotopy Classification of Maps. Lecture Notes in Mathematics, Vol. 628. Springer-Verlag, 1977.

Bednarek, C. and L. Cesari (Eds.): Dynamical Systems. Proceedings of the University of Florida International Symposium, 1976. Academic Press, 1977.

Bienayme, I.J.: Statistical Theory Anticipated. Studies in the History of Mathematics and Physical Sciences, Vol. 3. Springer-Verlag, 1977.

Bierstedt, K. and B. Fuchssteiner (Eds.): Functional Analysis: Surveys and Recent Results. Proceedings of the Paderborn (Federal Republic of Germany) Conference, 1976. Elsevier North-Holland, 1977.

Boutot, J.-F.: Schéma de Picard Local. Lecture Notes in Mathematics, Vol. 632. Springer-Verlag, 1978.

Chang, C.C. and H.J. Keisler: Model Theory, Second Edition. Elsevier North-Holland, 1977.

Chein, O.: Moufang Loops of Small Order. Memoirs, Vol. 197. American Mathematical Society, 1978.

Coleff, N.R. and M.E. Herrera: Les Courants Résiduels Associés à Une Forme Méromorphe. Lecture Notes in Mathematics, Vol. 633. Springer-Verlag, 1978.

Cooper, J.B.: Saks Spaces and Applications to Functional Analysis. Elsevier North-Holland, 1978.

Coppel, W.A.: Dichotomies in Stability Theory. Lecture Notes in Mathematics, Vol. 629. Springer-Verlag, 1978.

Csákány, B. and J. Schmidt (Eds.): Contributions to Universal Algebra. Colloquia Mathematica Societatis János Bolyai, Vol. 17. Elsevier North-Holland, 1977.

Curran, M.P.J. (Ed.): Topics in Group Theory and Computation. Proceedings of the Royal Irish Academy Summer School, University College, Galway, 1973. Academic Press, 1977.

Dixmier, J.: C*-Algebras. Elsevier North-Holland, 1977.

Erdelyi, I. and R. Lange: Spectral Decompositions on Banach Spaces. Lecture Notes in Mathematics, Vol. 623. Springer-Verlag, 1977.

Fiorini, S. and R.J. Wilson: Edge-colourings of Graphs. Pitman, 1977.

Gonzalez — Sprinberg, G.: Eventail en dimension 2 et transforme de Nash. Secretariat Math. de l'Ecole Normale Supérieure, 1977.

Grandy, R.E.: Advanced Logic for Applications. Reidel, 1977.

Grauert, H. and R. Remmert: Theorie der Steinschen Räume. Grundlehren der Mathematischen Wissenschaften, Vol. 227. Springer-Verlag, 1977.

Grenander, U. and L.H. Ballou: Pattern Analysis: Lectures in Pattern Theory, Volume 2. Applied Mathematical Sciences, Vol. 24. Springer-Verlag, 1978.

Guivarc'h, Y. M. Keane, and B. Roynette: Marches Aléatoires sur les Groupes de Lie. Lecture Notes in Mathematics, Vol. 624. Springer-Verlag, 1977.

Hammer, J.: Unsolved Problems Concerning Lattice Points. Pitman, 1977.

Harvey, W.J. (Ed.): Discrete Groups and Automorphic Functions. Proceedings of an Instructional Conference Organized by the London Mathematical Society and Cambridge University, 1975. Academic Press, 1977.

Heyer, H.: Probability Measures on Locally Compact Groups. Ergebnisse der Mathematik, Band 94. Springer-Verlag, 1977.

Kaplansky, I.: Hilbert's Problems. Preliminary edition. University of Chicago, Department of Mathematics, 1977.

Kauffman, R.M., T.T. Read, and A. Zettl: The Deficiency Index Problem for Powers of Ordinary Differential Expressions. Lecture Notes in Mathematics, Vol. 621. Springer-Verlag, 1977.

Kline, M.: Why the Professor Can't Teach: Mathematics and the Dilemma of University Education. St. Martin's Press, 1977.

Klingenberg, W.: Lectures on Closed Geodesics. Grundlehren der Mathematischen Wissenschaften, Band 230. Springer-Verlag, 1978.

Krishnaiah, P.R. (Ed.): Multivariate Analysis IV. Proceedings of the Fourth International Symposium, Wright State University, 1975. Elsevier North-Holland, 1977.

Lachlan, A., M. Srebrny, and A. Zarach: Set Theory and Hierachy Theory V. Bierutowice, Poland, 1976. Lecture Notes in Mathematics, Vol. 619. Springer-Verlag, 1977.

Landsberg, P.T. and D.A. Evans: Mathematical Cosmology. Oxford University Press, 1978.

Little, C. (Ed.): Combinatorial Mathematics V. Proceedings of the Fifth Australian Conference, Royal Melbourne Institute of Technology, 1976. Lecture Notes in Mathematics, Vol. 622. Springer-Verlag, 1977.

Mathematical Society of Japan and S. Iyanaga and Y. Kowada (Eds.): Encyclopedic Dictionary of Mathematics. Translated by the Mathematical Society of Japan with the Cooperation of the American Mathematical Society. MIT Press, 1977.

Matos, M.C. (Ed.): Infinite Dimensional Holomorphy and Applications. Elsevier North-Holland, 1977.

Nathanson, M.B. (Ed.): Number Theory Day. Proceedings of the Conference Held at Rockefeller University, New York, 1976. Lecture Notes in Mathematics, Vol. 626. Springer-Verlag, 1977.

Peking Computer Institute: Chinese-English Dictionary of Computer Terms. 1977.

Pfeffer, W.F.: Integrals and Measures, Marcel Dekker, 1977.

Pilz, G.: Near Rings: The Theory and Its Application. Elsevier North-Holland, 1977.

Popp, H.: Moduli Theory and Classification Theory of Algebraic Varieties. Lecture Notes in Mathematics, Vol. 620. Springer-Verlag, 1977.

Prolla, J.B.: Approximation of Vector Valued Functions. Elsevier North-Holland, 1977.

Robinson, A.: Complete Theories. Second edition. Elsevier North-Holland, 1977.

Sally, J.D.: Numbers of Generators of Ideals in Local Rings. Lecture Notes in Pure and Applied Mathematics, Vol. 35. Marcel Dekker, 1978.

Segal, I.E. and R.A. Kunze: Integrals and Operators. Second revised and enlarged edition. Grundlehren der Mathematischen Wissenschaften, Band 228. Springer-Verlag, 1978.

Serre, J.-P. and D.B. Zagier (Eds.): Modular Functions of One Variable VI. Proceedings of the International Conference, Bonn, 1976. Lecture Notes in Mathematics, Vol. 627. Springer-Verlag, 1977.

Smiley, T.J. and D.J. Shoesmith: Multiple - Conclusion Logic. Cambridge University Press, 1978.

Stoll, W.: Invariant Forms on Grassmann Manifolds. Annals of Mathematics Studies, 89. Princeton University Press, 1978.

Szabo, M.E.: Algebra of Proofs. Elsevier North-Holland, 1977.

Thue, A.: Selected Mathematical Papers. With Introduction by C.L. Siegel. Edited by T. Nagell. Universitetsforlaget, Oslo, 1977.

Watson, G.A.: Numerical Analysis. Proceedings of the Biennial Conference Held at Dundee, 1977. Lecture Notes in Mathematics, Vol. 630. Springer-Verlag, 1978.

Williams, N.H.: Combinatorial Set Theory. Elsevier North-Holland, 1977.

Zygmund, A.: Trigonometric Series. Reissued in one volume. Cambridge University Press, 1978.

*****

The Kalman Filter

A. V. Balakrishnan

The 'Kalman Filter' is best described as a soft-ware data-processor designed to filter signal from noise. It has been hailed as the single specific advance in the Systems area with significant impact on the battle front, according to at least one purported Air Force survey. Beginning with its application to satellite orbit tracking, it is now in routine use in all Navigation (position location) Systems: Inertial Navigation, Satellite Navigation, Radio Navigation as well as Radar and Sonar Tracking Systems. In military applications it is an integral part of Guidance Systems (where essentially the missile is made to follow a desired trajectory). In the near future it may well be employed in all automobile ignition systems.

For the mathematician it is just a sub-topic in statistics–stochastic processes – an area in which the revered mathematical names are Wiener and Kolmogorov. And yet it is remarkable that this particular theory did not originate from any traditionally trained Probabilist, etc., but rather from a 'system scientist' R. E. Kalman. [It is curious that there is a similar parallel in Information Theory – that after the pioneering work of Claude Shannon, despite the efforts of many bona-fide statisticians, by and large, the only significant (to communication engineering) result was that of A. Viterbi, a communication engineer who developed the decoding scheme bearing his name for convolutional codes.]

As part of the theory of stochastic processes, the Kalman Filter, while of immediate practical utility, can be described in 'axiomatic' mathematical language, and indeed rather sophisticated mathematics at that. In this respect, it is unlike the traditional view among mathematicians of "application" as mathematical physics, boundary value problems and associated numerical techniques. The general image that many mathematicians have of "Applied Mathematics" is basically of 'sticking numbers' into 'formulae'. The Kalman Filter is the best illustration of how wrong this view can be. It does not "solve" any equation and has nothing to do with physics. It does come up with a "formula" or "recipe" for processing data but 'application' does not mean merely putting numbers into formulae – indeed any real discussion of a Kalman Filtering application requires no less than a 'case study' full of the detailed and specialised jargon of the application area involved.

Central to the success of the Kalman Filter are two ideas: one is the exploitation of the 'dynamic system' representation of the signal and the other is the 'white-noise' representation of the noise. The early notion of a filter was simply one that 'cuts off' all frequencies beyond some value that did not distort the signal. Then came the theories associated with the names of Wiener and Kolmogorov. These theories were in a sense too sophisticated – the models as well as the results were too general to be of value in most applications. Indeed it is an oft repeated cliche in communication engineering that no Wiener filter was ever implemented.

It is necessary now to describe, however briefly, the mathematical theory involved so that we can be more precise. The Kalman Filter theory can be developed for 'continuous-time' or 'discrete-time' ('time-series') models. Since all Kalman filtering currently involves digital computation, the latter may be thought to be more appropriate but we shall see that this is not necessarily so. Since this note is addressed to mathematicians, we shall be concerned primarily with the basic assumptions and results rather than with the details of the derivation. We begin with the discrete-time theory in its most useful – rather than the most general – form.

Let $\{y_n\}$ denote the 'observed data' (sensor output, for example) sequence which is assumed to have the structure:

$$y_n = S_n + N_n$$

where $\{N_n\}$ is 'white noise': that is to say, N_n is a sequence of independent zero mean and known variance (and hence taken to be unit for simplicity) Gaussian variables, and $\{S_n\}$ is the "information-bearing" signal. Here $\{N_n\}$ represents the unavoidable 'instrument' error that remains even after all known sources of 'systematic' errors have been corrected for. An equally important assumption is that the sequence $\{S_n\}$ has the 'dynamic' representation

$$S_n = Cx_n$$
$$x_{n+1} = \Phi x_n + F\zeta_n$$

where C, Φ, F are known matrices and $\{\zeta_n\}$ is white noise again, and (for simplicity) statistically independent of the

noise sequence $\{N_n\}$. The Kalman Filter is designed to provide the 'least-squares' estimate of S_n based on data observed up to n: that is, to calculate

$$\hat{S}_n = E[S_n]y_m, \quad m \leqslant n]$$

The remarkable consequence of the assumed model (in contrast to the more general models of Wiener) is that we may express this estimate in a recursive way ('updating'):

$$\hat{S}_n = C\hat{x}_n$$

$$\hat{x}_{n+1} = \Phi\hat{x}_n + K_{n+1}[y_{n+1} - C\Phi\hat{x}_n]$$

Here $\hat{x}_n$ is the current estimate and the second term represents the "updating" based on the difference from the 'new piece of data' y_{n+1}. The advantage of the recursive estimate is that all the past data need not be stored. The sequence $\{K_n\}$ does not depend on the data and can be calculated "off-line", and moreover (under usual conditions) its asymptotic (large n) value is a calculable constant. Above all is the simplicity of structure of the filter.

The continuous-time version is similar but does involve a mathematical difficulty concerning the representation of the observation noise. Thus let

$$y(t)$$

represent the observed data and assume that

$$y(t) = S(t) + N(t)$$

where $S(t)$ is the signal characterised by the model:

$$S(t) = Cx(t)$$

$$\dot{x}(t) = Ax(t) + Fn(t)$$

where $n(t)$ is "white noise' of unit spectral density (matrix), A and F are given matrices. The observation noise $N(t)$ is also taken to be white independent of the process $n(t)$ and also of unit spectral density. Note that unlike the discrete case, the 'white-noise' processes are mathematically awkward because they have infinite variance. It is currently in vogue to use the 'integrated version':

$$Y(t) = \int_0^t y(s)\,ds = \int_0^t S(\sigma)\,d\sigma + W(t)$$

where $W(t)$ is a Wiener process, as being "more rigorous". This is nonsense because no physical instrument can output a Wiener process. The noise process is simply an idealization of the fact that the spectral bandwidth must be much larger than that of the signal and 'flat' over this range if the instrument is not to 'distort' the signal. More on this below. In any event, defining (with suitable interpretation):

$$\hat{S}(t) = E[S(t)\,|y(\sigma), \sigma \leqslant t]$$

$$\hat{x}(t) = E[x(t)\,|y(\sigma), \sigma \leqslant t]$$

we have

$$\hat{S}(t) = C\hat{x}(t)$$

$$\dot{\hat{x}}(t) = A\hat{x}(t) + P(t)C^*[y(t) - C\hat{x}(t)]$$

$$= (A - P(t)C^*C)\hat{x}(t) + P(t)C^*y(t)$$

the second form being particularly important in the asymptotic version ($t \to \infty$) when $P(t)$ is a constant.

There is a dilemma faced by every knowledgeable designer as to which model to use – discrete or continuous. If we assume the discrete model we must sample the data slowly enough to insure that the noise samples are indeed independent, but only the noise r.m.s. value need be known. On the other hand if we assume the continuous model it implies that we may sample that data at as high a rate as we wish but now we must know the spectral density. Here is one illustration of the point made earlier – 'application' is no longer just putting numbers into formulae. Rather difficult and often subtle decisions have to be made concerning the model to be used! Then again there is often the equally difficult question: how does one make sure one is right? The usual way out is to use a "computer simulation" which actually proves little, being a tautologous procedure. It is interesting in this connection to mention that just as the mathematicians misconstrue the Kalman Filter because it uses mathematics, in the same way the engineering users (most of whom have only a dim understanding of the theory) tend to look upon the Kalman Filter as a magic black box – and think that 'twiddling the dials' enough they can make it "work" – in spite of the theoreticians!

Of course in all the applications the 'linearity' of the equations is only an approximation and the instrument noise is not really white (or even Gaussian). The filter is "robust" enough to be good in spite of these limitations within reasonable bounds. What these bounds are – the judgment between too complex (and intractable) a model and too simple (and hence inadequate) is what the

designer's ingenuity must settle. No mathematician who has not had his "hands dirty" with a real, practical problem of this sort can appreciate what is involved – what "systems" work means – and herein appears again the difference from the usual "boundary-value problem numerical analysis" view of applications. This is of course not to say that significant – even great – contributions cannot come from non-practitioners.

The "brittleness" of the noise model is made more apparent in the push toward extending this theory to 'non-linear' dynamic models for the signal. Indeed the Wiener process models lead not only to contradictions but also to formulas which are meaningless – because the assumption that the 'integrated version' contains a Wiener process is simply the wrong extension of the model! However this cannot be appreciated by the "Non-Linear Filter" theorists who have never bothered to use any of their results on any 'real' data. Of course the fact that much of the current theory uses models which are not appropriate is not necessarily an argument against it!

References

The basic literature:

1. Kalman, R. E.: *A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, Vol. 1, 1960, p. 597–622. (where it all began)*
2. Battin, R.H.: Astronautical Guidance, McGraw Hill, 1964. (Space Guidance Application)
3. Bucy, R. S. and Joseph, P. D.: *Filtering for Stochastic Processes with Applications to Guidance, Interscience N.Y., 1968. (Source Book)*
4. Meditch, J. S.: *Stochastic Optimal Linear Filtering and Control, McGraw Hill, 1969. (Discrete Case)*

For a detailed account of the Wiener Process Version of Linear and Non-Linear Filtering:

1. Liptser, R. S. and Shiryayev, A. N.: *Statistics of Random Processes, Vol. I, and II – Springer-Verlag 1977.*

Acknowledgment

My warm thanks to Professors R. E. Kalman and R. E. Mortensen who were kind enough to read and criticise an early draft. The views expressed, needless to say, are my own.

System Science Department
University of California
Los Angeles, USA

New Undergraduate Texts

F.H. CROOM
Basic Concepts of Algebraic Topology
1978. 46 figures. X, 177 pages
DM 36,–; US $ 18.00
ISBN 3-540-90288-0

Basic Concepts of Algebraic Topology is designed for use as an introductory text at the undergraduate and beginning graduate levels. It emphasizes a geometric approach and may be considered a reaction against the high-powered axiomatic method of teaching algebraic topology. Concepts are illustrated with simple examples, and the proofs originally given by the discoverers of the important theorems are used when they are consistent with the introductory level of the course.

E.J. LECUYER, jr.
Introduction to College Mathematics with A Programming Language
1978. 126 figures. 64 diagrams. XII, 420 pages
Cloth DM 34,50; US $ 17.30
ISBN 3-540-90280-5

This innovative text contains the topics usually covered in a full years' course in finite mathematics or mathematics for liberal arts students. The exposition of the mathematical material is accompanied by an explanation of the computer language APL which then, in turn, is used for the treatment of applications and, in general, for a deeper understanding of the theory. The many problems accompanying the text lead the student to a more thorough understanding.

L. SMITH
Linear Algebra
1978. 21 figures. VII, 280 pages
Cloth DM 27,–; US $ 13.50
ISBN 3-540-90235-X

This book contains a sophomore level course in linear algebra and presents the basic, essential facts, as they will be needed to be built upon in further mathematics courses. Its goal is the derivation and proof of the Principal Axes Theorem for symmetric linear transformations. There is a wealth of numerical examples which illustrate the more abstract parts of the theory. More than 200 exercises are included.
Prices are subject to change without notice

The Birkhoff Prize Talks Atlanta, 1978

Following are the talks given by Professors Birkhoff, Kac, and Truesdell on the occasion of their acceptance of the Birkhoff prizes at the January meeting of the American Mathematical Society. The George David Birkhoff Prize in Applied Mathematics was established in 1967 in honor of Professor George David Birkhoff. The initial endowment of $2,066 was contributed by the Birkhoff family. It is normally awarded every five years, beginning in 1968, for an outstanding contribution to "applied mathematics in the highest and broadest sense." The award is made jointly by the American Mathematical Society and the Society for Industrial and Applied Mathematics. The recipient must be a member of one of these societies and a resident of the United States, Canada, or Mexico.

First award, 1968: To Jürgen K. Moser for his contributions to the theory of Hamiltonian dynamical systems, especially his proof of the stability of periodic solutions of Hamiltonian systems having two degrees of freedom and his specific applications of the ideas in connection with this work.

Second award, 1973: To Fritz John for his outstanding work in partial differential equations, in numerical analysis, and, particularly, in nonlinear elasticity theory; the latter work has led to his study of quasi-isometric mappings as well as functions of bounded mean oscillation, which have had impact in other areas of analysis.

Third award, 1973: To James B. Serrin for his fundamental contributions to the theory of nonlinear partial differential equations, especially his work on existence and regularity theory for nonlinear elliptic equations, and applications of his work to the theory of minimal surfaces in higher dimensions.

Fourth award, 1978: To Garrett Birkhoff for bringing the methods of algebra and the highest standards of mathematics to scientific applications.

Fifth award, 1978: To Mark Kac for his important contributions to statistical mechanics and to probability theory and its applications.

Sixth award, 1978: To Clifford A. Truesdell for his outstanding contributions to our understanding of the subjects of rational mechanics and nonlinear materials, for his efforts to give precise mathematical formulation to these classical subjects, for his many contributions to applied mathematics in the fields of acoustic theory, kinetic theory, nonlinear elastic theory, and the thermodynamics of mixtures, and for his major work in the history of mechanics. (Information courtesy of the Notices of the American Mathematical Society.)
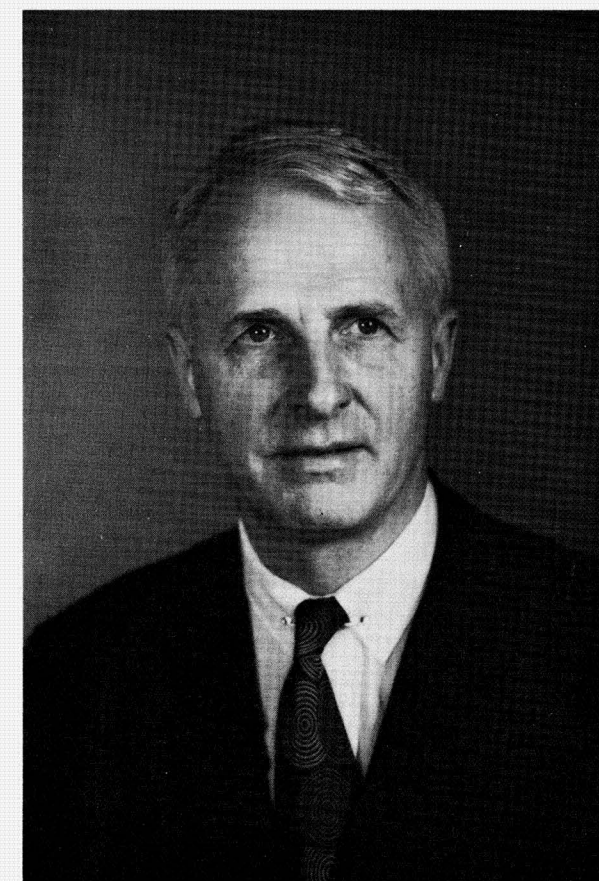
Garrett Birkhoff

I am naturally gratified sentimentally as well as scientifically by being awarded a George D. Birkhoff Prize. To express my pleasure, I shall try to describe briefly what I consider to be my most interesting algebraic contributions to Applied Mathematics.

1. *Applications of lattice theory.* From the 1930's on, I have been looking for new applications of lattice and group theory to scientific problems. I shall discuss

these applications first; their concern with existence and uniqueness theorems and with general qualitative properties comes closest in spirit to my father's own work. Perhaps one of my original aims was to convince him that "modern" algebra, in spite of its abstractness, was relevant to the real world!

In this connection, my 1936 paper [1] with von Neumann on the logic of quantum mechanics should be mentioned first. Our thesis was that the logic of quantum mechanics, like the lattice $L(H)$ of all closed subspaces of (separable) Hilbert space H, is at least a weakly modular ortholattice. Our paper stimulated much further research, which I reviewed (through 1960) in [2, pp. 156-63]. My only regret is that, because propositions about the position and momentum of a free particle apparently generate a *non*modular sublattice, this propositional calculus is not orthomodular!

A more sustained series of applications of lattice theory was inspired by the ergodic theorem. This asserts that, for any *measure-preserving* operator T acting on a probability space, the sequence of averages $(f + Tf + T^2f + ... + T^{n-1}f)/n = g_n$ *converges* to a fixpoint g of T. In 1938, I showed that this remained true in the sense of weak convergence even if T is not deterministic but only stochastic [3]. Soon afterwards, S. Kakutani and F. Riesz showed that the convergence was actually strong (metric); see [2, pp. 173-8].

A much more extensive series of applications is suggested by the work of Perron and Frobenius on nonnegative matrices; it applies to (female) population models, economic growth models, and nuclear reactors alike. The result is that a wide class of positive linear operators T have a positive eigenvector ϕ with positive eigenvalue α the ("multiplication factor") and a positive adjoint "importance" vector ψ, such that for any f, $T^nf \sim \alpha^n(f,\psi) \phi$. In female populations, ψ is the fertility functional and α the "growth factor" per generation; for neutrons, birth is caused by fission, death by absorption, and migration by diffusion.

In 1957, I gave the first general constructive proof of this result,[1] based on the discovery that T represented a contraction in a projective quasi-metric invented by Hilbert. I also pointed out that this theorem explained the existence and stability of critical distributions in nuclear reactors; see [2, pp. 178-84] and [4, pp. 116-26]. In 1962, I extended the result from semigroups to a wide class of "uniformly semiprimitive" multiplicative processes [5].

My review of applications of lattice theory in [2, pp. 150-84] also contains an exposition of the algebra of Reynolds operators, which axiomatizes (following original ideas of Kampé de Fériet) the properties of averaging operators in turbulence theory. I probably should have, but did not review in [2], my 1956 proof with Diaz (since extended in collaboration with Bruce Kellogg) of the existence and uniqueness of flows in networks with nonlinear resistance [6].

2. *Groups and fluid mechanics.* In the late 1940's, I also made a thorough analysis of the *Lie* groups arising in fluid mechanics. This analysis is closely related to the notable work of the Russian mathematician L. I. Sedov [7], which began some years earlier. We were both concerned with explaining the successful use of small-scale models in hydraulics, towing tanks, windtunnels, etc. Using *dimensional analysis* (high-school algebra and physics), Froude, Reynolds and others had predicted the mathematical *similarity* of suitably scaled flows past such models with those past full-scale ships, airplanes, etc. Sedov and I used *inspectional analysis* to correlate their reasoning with the partial differential equations of fluid motion. I suggested using Freon at reduced pressure to model inertial effects as well as cavitation [8].

Our second aim was to apply the *similarity* concept to solve partial differential equations involving n independent variables by functions of fewer variables (often only one). My conclusions concerning these problems are explained in Chapters 4 and 5 of my book on hydrodynamics [8].

I am sorry to say that Sedov and I seldom found the deeper parts of Lie theory useful for our work. However, these do seem to be relevant for its important recent extensions by Ovsjannikov [9].

Among my applications of Lie theory to fluid mechanics, most interesting by far from a purely mathematical standpoint is my geometrization of the Kelvin-Kirchhoff theory (see Lamb, *Hydrodynamics,* Ch. VI) of solids moving in an ideal fluid [8, § 112]. The motion of such a solid defines (relative to its original position) a trajectory on the Euclidean group manifold E. Relative to its added mass or *inertial energy* tensor T_{ij}, E becomes a Riemannian manifold with distance $ds^2 = \frac{1}{2} \Sigma T_{ij}dx^idx^j$. I showed how to express the (theoretical) fluid force acting on the solid in terms of the geodesic curvature of the trajectory and the structure constants of the Euclidean group.

3. *Practical applications; shaped charges.* Although not always requiring deep mathematical methods, most of the contributions I have mentioned so far are essentially to *natural philosophy*, rather than "practical".[2] Their purpose was to systematize and *rigorize* mathematical formulations of problems of pure science, in the spirit of Poincaré and G. D. Birkhoff's ergodic theorem.

Practical applied mathematics has a very difficult spirit, brilliantly described by Thornton Fry [10] in 1940,[3] when our country was beginning to "tool up" for the war against Hitlerism. Fry pointed out that in successful industrial (or military) mathematics, an appreciation of non-mathematical considerations is much more essential than mathematical sophistication.

My first contribution to practical mathematics was to explain the effectiveness of shaped charges, which were a major antitank weapon in World War II. Their theory, proposed independently by G. I. Taylor and myself (see [12]) uses only high-school algebra, geometry, and physical conservation laws of mass, momentum, and energy. We were led to it by recognizing the classic pattern of impinging jets in the first X-ray shadowgraphs of imploded conical steel liners; for simplicity, we *approximated steel by an ideal fluid,* as postulated by Kelvin and Kirchhoff. Crude and simple as our theory was, it was helpful to designers of shaped charges at a critical time in history.

4. *Scientific computing.* Since 1950, I have been increasingly concerned with practical applied mathematics, and especially with large-scale scientific computing. I had concluded by 1945 that the main *practical* obstacle to solving many elliptic boundary value problems was not analytical but *algebraic* — actually, *arithmetic!* Specifically, the problem was to solve *economically* Ax = b, where A was a large (1000 x 1000, say) "sparse" matrix representing an elliptic difference operator. This is a many-faceted problem of *real linear algebra,* far removed from group and lattice theory!

Too busy to tacke it myself, I had the good fortune to supervise David Young's 1950 Ph.D. Thesis [13], in which he derived and provided a rigorous theoretical explanation for his now classic SOR method for computerizing the solution of self-adjoint linear elliptic boundary value problems.

Young's thesis was published in 1954, and during the next several years I acted as consultant to Westinghouse, where Richard Varga was designing codes adapting Young's SOR method to the multigroup diffusion equations of nuclear reactor theory. The same differential equations arising in petroleum reservoirs were solved by a competitive ADI method invented a few years later by Peaceman, Rachford and Douglas. Young, Varga and I collaborated on a long joint paper [14] comparing these methods. Varga and I also wrote a paper [15] applying the theory of positive linear operators to the multigroup diffusion equations. But most valuable to me was the insight which my contact with Westinghouse gave me into industrial computing, and especially into the mathematics of nuclear reactor design.

During the 1950's, the first decade of large-scale computing, I was also associated as a consultant with the development at Los Alamos of early algorithms for computing the motions of fluids. Some of the work I did (on stationary flows with free boundaries) is reviewed in [16, pp. 117-140] and [17, pp. 55-76]. Although computers at that time still had very limited capabilities, I became increasingly convinced of the importance of *arithmetizing analysis in a practical sense,* with the help of high-speed computers. This would complement the great achievement of the 19th century, in which Analysis was "arithmetized" in a philosophical sense.

I have written up my ideas about solving economically (by computer) elliptic boundary value problems and fluid flow problems in two monographs ([18] and [19]), to which I refer you. I hope that clearly documented packages of computer programs for doing both, named ELLPACK AND WAVPACK respectively, will be generally available and easy to use within a few more years.

5. *Surface fitting.* It is much easier to *arithmetize the geometry of smooth surfaces,* and indeed the representation of sheet metal surfaces was successfully computerized in 1960-62 at the General Motors Research Laboratories, in cooperation with their Manufacturing and Development Staff. Carl de Boor and I explained the underlying mathematical ideas publicly a few years later [20, pp. 164-90].

The trick is to approximate a given surface $z = \phi(x,y)$ by smooth *piecewise polynomial* functions. For example, as proposed in 1961 by Carl de Boor,[4] $\phi(x,y)$ can be approximated by a *bicubic spline* function

$$F(x,y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{i,j}^{k,l} x^i y^j$$

on $[x_{k-1}, x_k] \times [y_{l-1}, y_l]$

constraining the $a_{i,j}^{k,l}$ to make F twice-differentiable. The space of all such functions, for a given product partition $\pi \times \pi'$ of a rectangle $[x_0, x_m] \times [y_0, y_n]$, is the *tensor product* of spaces consisting of the *univariate* cubic splines invented in 1742 by D. Bernoulli and Euler to describe the deflection of point-loaded thin beams.

6. *Interpolation algebra.* Applications should interest even the purest mathematicians, if only because they provide such a rich and suggestive source of mathematical problems. Thus many fascinating algebraic problems concern automata and coding [22, pp. 1-20].

---

[1]Krein and Rutman had previously proved it by compactness arguments.

[2]Newton emphasized the difference in the Preface to the first edition of his *Principia,* and explicitly disclaimed being "practical."

[3]I have published in [11] my own assessment of that nebulous area "applied mathematics".

[4]In 1960, H. L. Garabedian and I had proposed a similar but less simple scheme (J. Math. and Phys. 39 (1960), 258-68).

Likewise, at about the same time that the automobile industry was computerizing surface representation, piecewise polynomial functions became widely used in structural mechanics [21, pp. 210–42], and the name "finite element method" was coined to describe this use. Thanks in large part to these applications, and to the new problems they have suggested, interpolation and approximation have again become lively subjects, after being considered moribund for decades.

In particular, interest has revived in the so-called *Birkhoff-Hermite interpolation problem*. In a paper written when he was only 20, George D. Birkhoff asked for which set of n + 1 pairs of nonnegative integers $(k_i, x_{I(i)})$, is it true that there always exists a polynomial f(x) of degree n satisfying $f^{(k_i)}(x_{I(i)}) = c_i$ for any given set of c_i. This turns out to be a very difficult problem, solved by Polya in 1938 for the case of two points; G. G. Lorentz is writing a monograph on it.

On this occasion, it seems fitting to conclude by calling attention to an enormous generalization of this problem, which I first proposed in [22, pp. 38–45]. A special case concerns bicubic spline interpolation. The general problem is this: Given a polyhedral complex (in the sense of Poincaré), and values or compatibility conditions on its cells, what is the dimension of the space of piecewise polynomial functions of specified degree (possibly depending on the cell) satisfying *all* the given conditions?

As one small sample result, I mention the following. Let f be a continuous function, given by a polynomial function on each of the six faces of the cube $[-h, h]^3$. Then there exists a unique polynomial u(x,y,z) satisfying $u_{xxyyzz} = o$ interpolating to these boundary values; any sufficiently differentiable interpolant differs from f by $O(h^6)$; and any polynomial interpolant differs from f by a multiple of $(h^2 - x^2)(h^2 - y^2)(h^2 - z^2)$. This fact depends dramatically on the geometry of the cube: for the tetrahedron with vertices (o,o,o), (h,o,o), (o,h,o), (o,o,h), one only gets $O(h^4)$ accuracy, while in the sphere $x^2 + y^2 + z^2 = h^2$, since $h^2 - (x^2 + y^2 + z^2)$ vanishes identically on the boundary, the accuracy is only $O(h^2)$.

This example illustrates the enormously rich variety of possibilities on a general polyhedral complex, and I hope some of those present will be intrigued to think about them.

REFERENCES

[1] G. Birkhoff and J. von Neumann, "The logic of quantum mechanics," Annals of Math. 37 (1936), 823–43.

[2] R. P. Dilworth, editor, *Lattice Theory,* Proc. AMS Symposia in Pure Mathematics 2 (1960), American Mathematical Society.

[3] G. Birkhoff, "Dependent probabilities and spaces (L)," Proc. Nat. Acad. Sci. 24 (1938), 134–9.

[4] G. Birkhoff and Eugene P. Wigner, editors, *Nuclear Reactor Theory,* Proc. Applied Math. Symp. XI, Am. Math. Soc., 1961.

[5] G. Birkhoff, "Uniformly semi-primitive multiplicative processes," Trans. Am. Math. Soc. 104 (1962), 37–51.

[6] G. Birkhoff and J. B. Diaz, "Non-linear network problems," Quar. Appl. Math. 13 (1956), 431–43.

[7] L. I. Sedov, *Similarity and Dimensional Methods in Mechanics,* (first Russian Ed. in 1943); 4th ed. translated by Morris Friedman and Maurice Holt, Academic Press, 1959.

[8] G. Birkhoff, *Hydrodynamics: A Study in Logic, Fact and Similitude,* Princeton Univ. Press, 2nd ed., 1960.

[9] L. V. Ovsjannikov, *Gruppoye Svoystrva Differentialny Uravnenii,* Novosibirsk, 1962. Translation by G. W. Bluman, Univ. Calif. Los Angeles.

[10] Thornton C. Fry, "Industrial mathematics," Am. Math. Monthly 48 (1941). Special supplement (33 pp.).

[11] G. Birkhoff, "Applied mathematics and its future," pp. 82–105 of *Science and Technology in America: an Assessment,* Robb W. Thomson, editor, Nat. Bureau Standards Publ. 465, 1977.

[12] G. Birkhoff, D. P. MacDouglass, E. M. Pugh, and G. I. Taylor, "Explosives with lined cavities," J. Appl. Phys. 19 (1948), 363–82.

[13] D. M. Young, Jr., "Iterative methods for solving partial difference equations of elliptic type," Trans. Am. Math. Soc. 76 (1954), 92–111.

[14] G. Birkhoff, R. S. Varga, and D. M. Young, Jr., "Implicit alternating direction methods," Advances in Computers 3 (1963), 190–273.

[15] G. Birkhoff and R. S. Varga, "Reactor criticality and non-negative matrices," J. SIAM 6 (1958), 354–77.

[16] Monroe H. Martin, editor, *Fluid Dynamics,* Proc. Symposia Appl. Math. IV, Am. Math. Soc., 1953.

[17] G. Birkhoff, R. Bellman, and C. C. Lin, editors, *Hydrodynamic Instability,* Proc. Symposia Appl. Math. XIII, Am. Math. Soc., 1962.

[18] G. Birkhoff, *The Numerical Solution of Elliptic Equations,* Regional Conf. Series in Appl. Math. No. 1, SIAM Publ., 1971.

[19] G. Birkhoff and V. A. Dougalis, *Numerical Fluid Dynamics* (monograph to be published in 1979).

[20] H. L. Garabedian, editor, *Approximation of Functions,* Elsevier Press, 1965.

[21] G. Birkhoff and R. S. Varga, editors, *Numerical Solution of Field Problems in Continuum Physics,* SIAM-AMS Proceedings II, Am. Math. Soc., 1970.

[22] G. Birkhoff and Marshall Hall, Jr., editors, *Computers in Algebra and Number Theory,* SIAM-AMS Proceedings IV, Am. Math. Soc., 1971.

Vanderbilt University
Vanderbilt, Tennessee, USA
(On leave from Harvard University)
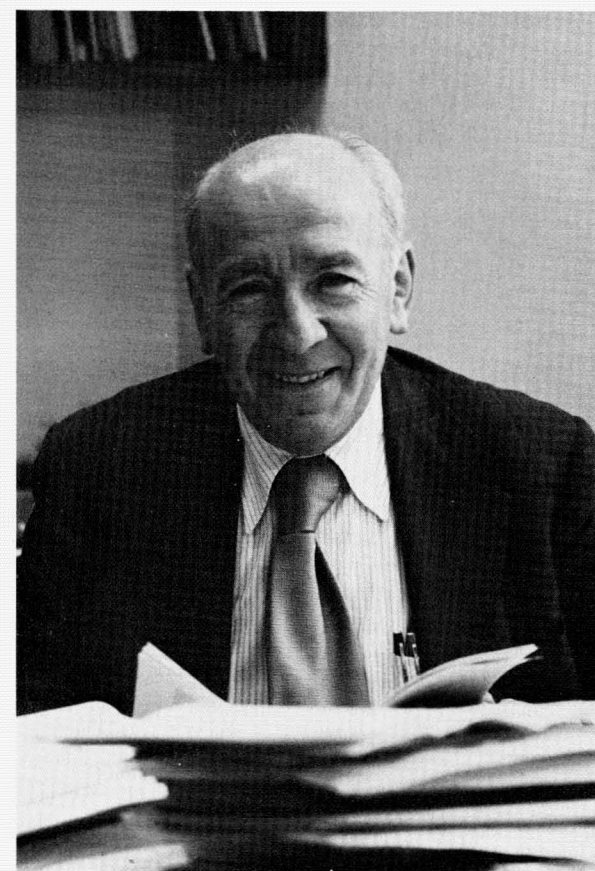
Mark Kac

I am greatly honored and naturally pleased at having been awarded the Birkhoff Prize. The letter notifying me of the award also invited me to say a few words at the presentation of the Prize but did not suggest any specific topic.

I find it somewhat awkward to speak of my own work and for a variety of reasons. One of them is that having been cited for contributions "to Probability Theory and its applications," I would feel compelled to admit that the application of Probability Theory, of which I am most fond, is the one which Paul Erdos and I made, nearly forty years ago, to Number Theory. Since the Prize is in Applied Mathematics and since Number Theory is not (yet!) a part of this domain, I could have found myself in (at least two-sided) jeopardy.

It was thus more appropriate (and safer) to say something about Applied Mathematics, but the lateness with which I was informed about being a recipient of the Prize and natural laziness conspired against preparing a special essay for the occasion. With apologies to all, I shall therefore read excerpts (with minor modifications) of a talk I gave in September 1971 at the Brown University Symposium on "The Future of Applied Mathematics" and published in a special issue of the Quarterly of Applied Mathematics in April, 1972:

I have felt for a long time that in discussing the perennial topic of the role of applied mathematics one is inclined to overlook that at the heart of the matter is the subtle and fundamental question of the relation of mathematics to non-mathematical disciplines.

We tend to think of "applied" as being roughly synonymous with useful or "practical," but I should like to argue that applying mathematics is an activity which often transcends pragmatic considerations and that we should also be concerned with a deeper exploration of what this activity is or ought to be.

As usual, it is easier to say what it ought not to be, and here I cannot resist referring once more to a wartime cartoon depicting two chemists surveying ruefully a little pile of sand amidst complex and impressive-looking apparatus. The caption read: "Nobody really wanted a dehydrated elephant but it is nice to see what can be done." I am sure that we can all agree that applying mathematics should not result in the creation of "dehydrated elephants."

But how can one tell?

A necessary (though by no means sufficient) prerequisite is to *know* the subject or discipline to which we apply mathematics and to know it with such intimacy as to understand its spirit and to appreciate its lore.

When we propose to apply mathematics we are stepping outside our own realm, and such a venture is not without dangers. For having stepped out, we must be prepared to be judged by standards not of our own making and to play games whose rules have been laid down with little or no consultation with us.

Of course, we don't have to play, but if we do we have to abide by the rules and above all not try to change them merely because we find them uncomfortable or restrictive.

From now on let me stick to physics, not only because it is the only non-mathematical discipline with which I am reasonably familiar but also because physics and mathematics have had close ties in the past and because after a period of alienation there are signs of a rapprochement. My joy at the prospect of the two great disciplines coming together again is somewhat mitigated by the fear of "dehydrated elephants" on one side facing a rigid and even hostile pragmatism on the other.

Since there is nothing I can do about pragmatism of physics (and some physicists!) let me limit my discussion to exploring some of the ways in which mathematics can be significantly applied to certain problems of physics. Let me also state that I speak only for myself and that I do not expect my views to be universally accepted. Quite to the contrary, I expect opposition, and would welcome a debate or even a confrontation.

First, it seems self-evident that mathematics is not likely to be much help in discovering laws of nature. If a mathematician wants to make a contribution on this (and I admit it is the highest) level, he will have to master so much experimental material and train himself to think in a way so different from the one he has been accustomed to that he will, in effect, cease to be a mathematician.* Perhaps it is well to be reminded by the way of analogy that while in recent years a number of physicists have made significant contributions to biology they accomplished this not because they *were physicists* but because they *became biologists*. I cannot, of course, claim that their *training* in physics had nothing to do with their successes in biology. More likely it has a great deal to do with it, and by the same token a mathematician who becomes a physicist (or a biologist, or an economist) may benefit greatly from his training as a mathematician, but this is not what I am concerned with today.

As *mathematicians* we come in when this or that law of nature has been discovered, and our role is usually twofold: (a) to help find ways in which the law can "best" be formulated and (b) to help in drawing conclusions, which hopefully will be significant to the further development of the subject. Of the two, the first role is fraught with more danger, since it tends to provide fertile ground for breeding "dehydrated elephants." This is especially true today because contemporary mathematics is geared to playing with "formulations" and physics, so full of vague, unprecise and unrigorous statements, seems to cry out for being better "formulated" and fitted into familiar structures.

Still, it would be narrow-minded and unjust to discourage searches for new formulations. No one could deny that Minkowski's formulation of the special theory of relativity has been decisive in the development of this theory, and it is beyond dispute that Riemannian geometry is the proper setting for Einstein's general theory. But one would find much less agreement as to the value of the subtle and elegant axiomatization of thermodynamics by Caratheodory, and one could cite examples of reformulations whose pertinence and value is even more in doubt.

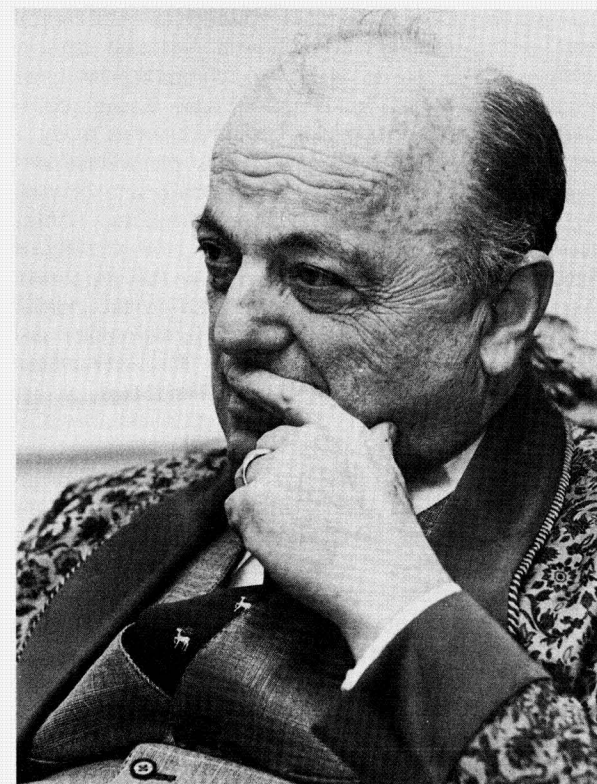It is not quite easy to explain why "geometrization" of relativity theory (both special and general) is among the most glorious applications of mathematics while Caratheodory's reformulation is a bit of a "dehydrated elephant," but it must have something to do with the distinction between facing up to problems which are, so to speak, God-given and those which are only man-made.

If, as it has been suggested, our role is to keep reexamining physics in the light of conceptual advances in mathematics, then without proper checks, this would, in my opinion, constitute the broadest license for creation of "dehydrated elephants." It would also add to a kind of self-delusion that one is contributing to a subject while in reality all one is accomplishing is discovering that all the time one has been speaking prose.

As a reliable check, I would propose testing formulations and reformulations against some of the questions (preferably hitherto unanswered) which this or that field evolved over a period of time. If a dent (even a small one!) can be made, the formulation has shown viability and should be pursued. If, however, all that comes out is the same old stuff in a different guise — well, write it off as self – education, put it in a notebook and above all don't feel compelled to publish it.

Within the last year or so (that is, roughly five years after I made the remarks I have just read) we have witnessed a new promising and exciting confluence of mathematical and physical ideas. I speak, of course, of modern Differential Geometry on the one hand and the theory of Gauge Fields on the other. Miraculous as it may seem, fibre bundles, homotopy, and Chern classes are becoming as much parts of physical terminology as instantons, gauge fields and Lagrangians are becoming parts of the mathematical one. Opportunities for fruitful collaboration and exchange of ideas abound, and a common language is being evolved. If this trend continues, the spectre of dehydrated elephants will vanish, and perhaps a future Birkhoff Prize will be awarded for work on homotopy classes of certain maps or for an extension of the index theorem.

---

*This runs counter these days to the view (professed by adherents of Catastrophe Theory) that "pure thought" alone can be a decisive instrument in comprehending Nature. This view (which goes back to Aristotle) has been propounded (notably by Descartes) in one form or another throughout history. That it surfaced once again in this day and age is a source of wonderment expecially if one recalls the fate of Aristotle's physics and Descartes' theory of blood flow.

*Department of Mathematics
Rockefeller University
New York, USA*

Clifford Truesdell

The honor you give me today is double: The first time you invite me to address you indebts me to you also for a splendid award. The occasion is twice doubly dear because it lets me say a little about two men closely connected with the prize: George D. Birkhoff himself, and a previous recipient, James Serrin. Both of these men have made an impress upon American and foreign styles in mathematics which I know to be healthful and hope will be lasting.

I met Birkhoff just once. It was at the summer meeting of the Society on September 12, 1943, at Rutgers. In those days there were about 2,000 members, of whom some 200 were active mathematicians; at a meeting in the East you could usually encounter about half of those. My teacher, Lefschetz, who was often kind to his students introduced me to Birkhoff and to many other senior men. After dinner there were two speeches. The report[1] published in the *Bulletin* is brief: "Dean R. G. D. Richardson of Brown University spoke of the importance of applied mathematics in the war effort. Professor G. D. Birkhoff of Harvard University urged mathematicians and scientists to maintain a proper balance of values during the emergency." Having survived a summer school of applied mathematics at Brown, I knew what to expect of Richardson; the *Bulletin* sums his lecture. Not so Birkhoff's. With commanding dignity and in superb, native English,

Birkhoff lashed the universities for their subservience to government and warfare. He warned us that by giving junior men heavy teaching loads, as much as eighteen hours, with no assistance, and by admitting unselected and mainly unqualified undergraduates, the universities were destroying what it was their prime duty, above nations and above emergencies, to foster. The many incompetents pressed into instruction were unable to teach, for they did not know, while the competent few were unable to learn because they were left no time.

I wonder what Birkhoff would have said, had he lived to see them, about our mass universities today, universities which in a time of peace and ease have forgotten that the task of higher learning is not only to sow, dung, and harvest but also and above all to winnow. I wonder what he would have said about the pollution by social sciences, changing values, team work, computing, sponsored research, involvement in the community, and even soft mathematics, to the point that mathematics of his kind is today decried as being "esoteric". I wonder what he would have said about the iron rule of mediocrity the Government today imposes as the price of the manna called "overhead" it scatters to the voracious and insatiable education mongeries. I wonder what he would have said about our management by regular corporate administrations which lack, however, the profit motive and in perfect Parkinsonian policy have but a single real interest alike in students and in employees (some of whom are still called "faculty"): Get money out of them directly or for them through subsidy, in quantity sufficient to sustain exponential growth in number and power of administrators — mad pursuit of bigness, until, one day, every boy and girl will be conscripted to serve a term of unstructured play garnished with social indoctrination called "higher learning", and all universities will be branch offices of one Government bureau.

After the meeting the next I heard of Birkhoff was that he had died. Like Birkhoff's speech, the beautiful description of him in the obituary[2] by Marston Morse is fixed so firmly in memory that I can almost quote it by heart:

….Birkhoff thought of his contemporaries in Europe, particularly in Germany, as colleagues rather than as teachers. He held Klein lightly, was unenthusiastic over Weierstrass, but gave his full respect to Riemann. Through his papers on non-self-adjoint boundary value problems and asymptotic representations he probably influenced the Hilbert integral – equation school as much or as little as it influenced him. His relations with the members of the French and Italian schools of analysis were close, both personally and scientifically. Levi-Civita and Hadamard were among his best friends. Birkhoff was at the same time internationally minded and pro-American. The sturdy individualism of

Dickson, E. H. Moore and Birkhoff was representative of American mathematics "coming of age." The work of these great Americans sometimes lacked external sophistication, but it more than made up for this in penetration, power and originality, and justified Birkhoff's appreciation of his countrymen.

Morse's words should dispel the idea, today often noised, that before the arrival of refugees from Hitler scant mathematical research and advanced teaching was done in this country. Of course, mathematics is above nations and nationalities, but that works both ways; moreover, much of the best teaching is done outside of the classroom and at great distance. Morse went on to say that "Poincaré was Birkhoff's true teacher."

Serrin, like Birkhoff, emerged from what Morse called "the dynamic individualism of the Middle West" and also experienced "the environment of tenacious self-sufficiency in New England. . . ." Like Birkhoff, Serrin was trained entirely in this country and largely in the Middle West. I met him when he was a student at the Graduate Institute for Applied Mathematics, Bloomington, Indiana. On September 30, 1950, just after I had arrived there, in my first letter to Charlotte Brudno, who was then my assistant and who later graced me by becoming my wife, I wrote: "As for students, at least one in my class is very good. . . . He has just succeeded in. . . . giving correct and relatively simple proofs of the theorems of Lavrentieff in free boundary theory." My next letter to her, two days later, opens with the words "Serrin is making brilliant discoveries. . . . Day by day new results appear. . . . If there is no error, this will be one of the two major achievements of our century in the theory of potential flow of incompressible fluids (the other being Levi-Civita's proof of existence of finite surface waves)." In this same class was Jerald Ericksen, also a Westerner by birth and training, a man who developed more slowly but in the end also has made contributions to the rational mechanics of solids and fluids that rank, it seems to me, second to none in our time.

The course was statistical mechanics. That was the first time I taught Birkhoff's ergodic theorem. We used Khinchin's book, available only in a stupid translation that forced us to create everything halfway afresh. On the final examination, Serrin outlined arguments which indicated his discovery of a formal connection between Hilbert's process and Enskog's process in the kinetic theory of gases. His grade was A+. It was the first time I gave that grade, and I cannot recall ever having given it since. Unfortunately Serrin never worked through his ideas on this central problem, which has since then been clarified by Grad and Muncaster, though still only formally.

When I survey, to the best of my limited ability, the splendid researches of Birkhoff and Serrin, I find several points of similarity. Both deserve Morse's words

of praise for the independent Americans: penetration, power, and originality. Both show a profound sense of problem, the one quality common to great mathematicians of all kinds and all ages. Technique, so prominent a feature in papers on pure analysis today, is there on demand, but it never takes precedence over concept. In neither's work is there a hard line between "pure" mathematics and its applications. About three-quarters of Birkhoff's subjects came directly or indirectly from mechanics; in Serrin's, the proportion seems reversed. Birkhoff was able also to face nature and knowledge as such and himself to formulate new theories for them. This part of his work deserves more study than it has received. Birkhoff endorsed, as had Helmholtz, the view that a mathematical theory of an aspect of nature is a model, and as such subject not only to mathematical development but also to scrutiny at its foundations. Above all we must not believe uncritically what physicists and engineers tell us. Morse, perhaps in allusion to the astonishing success of Einstein in the newspapers, wrote "Birkḥff inherited from Poincare the sentiment that no single mathematical theory of any phenomenon deserves the exclusive attention of physicists, or at least of mathematicians." Serrin also, in his new work on thermodynamics, faces the phenomena directly, selects the mathematical concepts in which to describe them, and lays down his own axioms to relate them. Theorems, real theorems — rare birds they have been in thermodynamics — come afterward, theorems that are not applications of known techniques but more likely the fathers of new methods in analysis.

Such is the tradition of Hilbert, who in the preface to his famous list of problems for the twentieth century to solve wrote

> While I insist upon rigor in proofs as a requirement for a perfect solution of a problem, I should like, on the other hand, to oppose the opinion that only the concepts of analysis, or even those of arithmetic alone, are susceptible of a fully rigorous treatment. This opinion. . . . . I consider entirely mistaken. Such a one-sided interpretation of the requirement of rigor would soon lead us to ignore all concepts that derive from geometry, mechanics, and physics, to shut off the flow of new material from the outside world. . . . What an important, vital nerve would be cut, were we to root out geometry and mathematical physics! On the contrary, I think that wherever mathematical ideas come up, whether from the theory of knowledge or in geometry, or from the theories of natural science, the task is set for mathematics to investigate the principles underlying these ideas and establish them upon a simple and complete system of axioms in such a way that in exactness

and in application to proof the new ideas shall be no whit inferior to the old arithmetical concepts.

The sixth of Hilbert's problems was "mathematical treatment of the axioms of physics", after the model of geometry. . . . "first of all, probability and mechanics." Hilbert's influence on recent rational mechanics is not so widely known as it should be. While he published scantly in that field, years ago I found in the libraries of Purdue University and the University of Illinois notes taken by two Mid-Westerners in Hilbert's course on continuum mechanics, 1906/7. After studying those notes, in my own expositions from 1952 onward I followed Hilbert's lead, not in detail but in standpoint and program; basic laws are formulated in terms of integrals, and the science is treated as a branch of pure mathematics, by systematic, rigorous development motivated by its logical and conceptual structure, decending from the general to the particular.

The word "axiom" may confuse, since on the one hand all of mathematics is essentially axiomatic, while on the other hand "axiomatics" is a term often used pejoratively to suggest a fruitless rigorization of what everybody believed already. A mathematician must know that good axioms spring from intrinsic need — details scatter, concept is not clear, or paradox stands unresolved. Formal axioms represent only one response to the call for conceptual analysis before significant new problems can be stated. You cannot solve a problem that has not yet been set. Formal axiomatics make an important part of modern rational mechanics; I refer in particular to Walter Noll's solution of the relevant portion of Hilbert's sixth problem. But by far the greater part of modern mechanics is neither more nor less axiomatic than is any other informally stated branch of mathematics. *The essence is conceptual analysis,* analysis not in the sense of the technical term but in the root meaning: logical criticism, dissection, and creative scrutiny. After that comes the poetry of statement and proof, the ornament of illuminating examples.

A mathematical discipline is made by mathematicians. However much they may begin from, digest, clarify, and build upon the ideas of physicists and engineers, it is the mathematicians who make sense out of them.

Now, I fear, many people accept a picture too narrow of mathematical activity. They forget that the great theories which enable us to understand in part the world about us and on which engineers and computers base their applications were created by great

mathematicians; they forget also that the problems suggested by new mathematical theories of nature often fail to fall within any of the previously existing fields of "applied mathematics", which fields, all of them, grew out of older theories of nature. They disregard a prophetic warning published in 1924 by that canny and shrewd observer, Richard Courant:

> . . . . many analysts are no longer aware that their science and physics. . . . belong to each other, while. . . . often physicists no longer understand the problems and methods of mathematicians, indeed, even their sphere of interest and their language. Obviously this trend threatens the whole of science: the danger is that the stream of scientific development may ramify, ooze away, and dry up.

I first read this warning in 1941, when I was a student of Bateman; it has been fixed in my memory ever since. I have chosen to mention Serrin's work as well as Birkhoff's because he was formed, not before but long after the danger had appeared, and he proves that mathematics of Birkhoff's kind is not impossible even today.

I should like to believe it this kind of mathematics — individual, self—sufficient, neither enslaved to nor isolated from natural science — that the Birkhoff prize is designed to recognize.

If conceptual analysis of physical problems is a part of mathematics, is it pure or is it applied? I think it is both. I believe that here I may count upon the agreement of not only Serrin, Courant, Birkhoff, Hilbert, and Poincaré[1] but also mathematicians of still higher rank: Cauchy and Euler.

This speech has been longer than it ought have been. There is an excuse. As my first invitation to address the Society has come thirty—six years after I joined it, I have tried to squeeze in everything I had to say, for fear that should a second arrive after the now established lapse of time, it would be not only long after 1984 but also long after the authorities would release me to accept the honor.

[1]*Bulletin of the American Mathematical Society* 43 (1943), 825.

[2]*Bulletin of the American Mathematical Society* 52 (1946), 357–391.

Department of Rational Mechanics
The Johns Hopkins University
Baltimore, Maryland, USA

Archimedes' Lost Treatise on the Centers of Gravity of Solids

W. Knorr

Introductory note. The present paper is one of a series developing out of my study of the chronological ordering of the Archimedean corpus ("Archimedes and the *Elements*," to appear late this year). This issue had been obscured by several implausible ideas introduced by Heiberg, reproduced by Heath (10, p. xxxii; 9, II, p. 22) and perpetuated uncritically in the secondary literature ever since. For instance, viewing the *Dimension of the Circle* as late and the *Method* as early, as Heath did, how could one avoid confusion on the development of Archimedes' work? In discovering Heath's error and reversing his judgment on these two works, I hit upon a criterion by which to divide the writings in the corpus into an "early group" and a "mature group." The former, including *Dimension of the Circle* and parts of *Quadrature of the Parabola,* apply techniques characteristic of Euclid's *Elements:* most notably, convergence is one-sided, a magnitude being approximated by a sequence of inscribed rectilinear figures. By contrast, the mature treatises invariably apply two-sided convergence, in which the magnitude is compressed between sequences of circumscribed and inscribed figures whose difference can be made arbitrarily small. Further, these works frequently employ the "Archimedean axiom" on convergence, instead of the Euclidean bisection principle (Euclid's *Elements X, I*).

Through this criterion the placement of the mechanical writings becomes clearer. For instance, by virtue both of their content and of their use of the Euclidean convergence technique, the two books *On Plane Equilibria* (establishing the principle of equilibrium and the centers of gravity of the parallelogram, the triangle, the trapezium and the parabolic segment) can be associated with the "early group." As said above, the *Method* ought to be set near the end of the corpus, so to be viewed as a retrospective on the "mechanical method" by which Archimedes had initially studied the problems presented in the mature treatises. Indeed, the theorems in the mature writings frequently betray signs of their preliminary investigation through the mechanical manipulation of indivisibles. The two books *On Floating Bodies* also must be viewed as late; for the second book requires the volume of segments of paraboloids, as given in *Conoids and Spheroids,* a work known to be the latest of the extant geometric treatises.

This revised chronology can form the basis for addressing questions on the development of Archimedes' techniques such as these: what were the methods and motives by which Archimedes advanced beyond Euclid's *Elements?* what did the mechanical procedures contribute, not only to the heuristic study of the problems, but also to the elaboration of formal techniques? how did the scholastic formalism of the geometers at Alexandria influence Archimedes' style of proof? Moreover, I have been led to the view that Archimedes depended on certain *pre*-Euclidean works, not the Euclidean edition of the *Elements,* as his source for proportion theory and other materials. I argue further that Pappus' alternative treatments of the sphere and the spirals (*Collection* IV and V), cast according to the techniques of the "early group," ought to be accepted as close paraphrases of early Archimedean versions, no longer extant.

I propose now to examine Archimedes' study of the centers of gravity of solids. His treatise on this subject, now lost, can be placed within the "mature group," following *Conoids and Spheroids,* but preceding *Floating Bodies* and the *Method.* From clues preserved in these three extant works I shall seek to identify the theorems it contained and the manner of their proof, and then to provide in outline two of them, on the centers of gravity of the cone and the paraboloid.

Renaissance commentators and translators of Archimedes, such as Maurolico and Commandino, remarked with surprise that Archimedes left no writings on the centers of gravity of solids. In a number of treatises they and others were thus stimulated to investigate these problems.[1] But manuscripts unavailable to them have since been recovered which establish that Archimedes had indeed worked out the solutions to such problems.[2] For instance, in the *Method* there appear the heuristic treatments of the center of gravity of the segment of the paraboloid and other solids of revolution. Moreover, in the second proposition of *Floating Bodies II* Archimedes remarks that this very result "has been proved in the *Equilibria* (ἐν ταῖς Ἰσορροπίαις)."[3] The statement of the theorem here is close to that of the fifth proposition of the *Method,*

save that the former enunciates the more general case where the plane cutting off the segment need not be perpendicular to the axis of the paraboloid.

Thus, we learn from Archimedes himself of a work, now lost, in which he had investigated the centers of gravity of solids. Moreover, his use of the word "prove" indicates that this work was formal in nature; for in the preface to the *Method* he makes clear the distinction between geometric demonstrations and heuristic procedures. To the same effect, in his formal works Archimedes dispenses with the proofs of needed theorems only when these had already been given, as in earlier treatises. The use of the center of gravity of the paraboloid in *FB II*, 2 − 10 thus points to a prior *formal* proof of this result.

In the present discussion we shall propose a reconstruction of the formal proof along the lines revealed in the *Method.* In this way the character of the lost treatise will emerge, so that a comparison with the methods adopted by the Renaissance scholars can be made.

First, let us review the argument in *Method,* prop. 5. Archimedes' general approach in this work is to consider the infinitesimal sections which constitute a magnitude — for instance, the circular sections of a solid of revolution which "fill out" its volume. Two such solids may be conceived as extended along the arms of a balance such that, taken in pairs, the corresponding sections of these solids are in equilibrium; from this it follows that the entire solids will likewise be in equilibrium. Thus, from the principle of the balance, one may determine the volumes if the centers of gravity are known, or, conversely, the centers if the volumes are known. In *M*, 5 the latter procedure is adopted.

Figure 1

Archimedes proposes to show how he discovered that the center of gravity of the solid generated by the revolution of the parabolic segment *ABC* about its axis *AD* (Fig. 1) is the point which divides the axis in the ratio 2:1 (the longer segment lying toward the vertex *A*). He introduces the cone generated by the revolution of the triangle *ABC*, extends the axis to *F* such that *DA = AF*, and imagines *DF* to be a balance with fulcrum at *A*. The solids are next cut arbitrarily by a plane perpendicular to the axis, producing in the paraboloid a circular section of diameter *OX*, and in the

cone a circular section of diameter *RP*. From the definition of the parabolic curve it is known that $DA : AS = BD^2 : XS^2$. After a few steps he obtains $FA : AS = XS^2 : SP^2$.*

Since circles have the ratio of the squares of their diameters, this result means that the circle of radius *XS* remaining centered at *S* is in equilibrium with the circle of radius *PS* centered at *F*. This relation will hold for every pair of circles formed by planar sectioning of the solids. Thus, when the solids are "filled out" by the circles, "all the circles in the segment of the paraboloid" in position along *AD* will be in equilibrium with all the circles of the cone so transferred that its center of gravity is at *F*. If *K* is the center of gravity of the paraboloid, then $FA : AK$ = (volume of paraboloid) : (volume of cone). The latter ratio is known to be 3 : 2 from the previous theorem (*M*, 4). Thus, since *FA = AD*, Archimedes can conclude that $AK : KD = 2 : 1$.[4]

This procedure may be adapted as the basis of a formal demonstration. How this is to be done emerges from a comparison of certain theorems in the *Method* with surviving formal versions — namely, *M*, 1 (on the area of the parabolic segment) with *Quadrature of the Parabola*, 14−16 and *M*, 4 (on the volume of the paraboloid) with *Conoids and Spheroids*, 21−22. The formalization in these cases entails, first, the replacement of infinitesimal elements by figures homogeneous with those to be measured and, second, the elimination of the physical conception of the balance.[4a] The proof itself must be cast as an indirect argument, establishing the claimed equality by developing a contradiction from the assumption of inequality. The theorem on center of gravity will differ from those in that the notion of the balance must be retained, for this is the basis of the concept of center of gravity; whereas, according to Greek formal criteria, mechanical principles are dispensable — and, hence, inappropriate — within a purely geometric problem, like the measurement of volume, they are necessary within a study such as this on centers of gravity.

Figure 2

Let the parabolic segment *ABC* revolve on its axis *AD*, producing the paraboloid, while at the same time the triangle *ABC* revolves to produce a cone (Fig. 2). (We

*By similar triangles, $DA : AS = DB : SP$. Thus, $DB : SP = BD^2 : XS^2$, or the lines *BD*, *XS*, *SP* are in geometric proportion; that is, $(XS : SP)^2 = BD : SP$. Since *DA = FA*, it follows that $XS^2 : SP^2 = FA : AS$, as claimed.

assume the diameter is perpendicular to the base.) By means of planes perpendicular to the axis, let each solid be sectioned, and let cylinders whose bases are these sections and whose axes lie along AD be marked off, so that each solid is bounded above and below by an ensemble of them. (We will take their heights to be equal.) Let P_n denote the figure consisting of n-many cylinders inscribed in the paraboloid, C_n the analogous figure inscribed in the cone, and P_n' and C_n' the corresponding circumscribed figures. As in the *Method*, we may compare the sections to find that $AD : AS = XS^2 : PS^2$. But if the cylinder whose base is the circle of radius PS is suspended at F (where $DA = AF$), it will be slightly too small to balance the cylinder whose base is the circle of radius XS, positioned where it is; for the center of gravity of the latter cylinder is at the midpoint of its axis, hence further from A than S. Denoting by K_n the center of gravity of the figure inscribed in the paraboloid, we thus obtain an inequality when all the cylinders are summed: $AK_n : AF > C_n : P_n$. In similar fashion, if K_n' denotes the center of gravity of the circumscribed figure, then $AK_n' : AF < C_n' : P_n'$. Now, by means of the summations which Archimedes develops in lemmas before the first and second propositions of *Conoids and Spheroids*, it follows that $C_n : P_n = 2n - 1 : 3n$ and $C_n' : P_n' = 2n + 1 : 3n$. Denote by L_n the point on the axis AD such that $AL_n : AD = 2n - 1 : 3n$; by L_n' the point such that $AL_n' : AD = 2n + 1 : 3n$; and by L the point such that $AL : AD = 2 : 3$. Then, we claim that K, the center of gravity of the paraboloid, is the point L.

That the bounding ensembles converge in volume to the paraboloid and the cone follows from CS, 21 ; as $P_n' - P_n$ and $C_n' - C_n$ each equal the single cylinder whose base is the circle of radius BD, these differences can be made arbitrarily small by marking off cylinders of suitably small height – that is, by taking n suitably large. As the centers of gravity of all the cylinders lie on AD, it follows that K_n and K_n' do also; just as Archimedes shows that the center of gravity of the parabolic segment lies on its diameter (*PE II*, 4), so one may show that K lies on AD.[5] By analogy with theorems on the parabola, one may further establish that $AK_n > AK > AK_n'$ (cf. *PE II*, 5) and that K_nK_n' can be made arbitrarily small (cf. *PE II*, 6).[6]

To prove $K = L$, we assume the contrary. First, let $AK < AL$. The segments AK_n are all greater than AK, while the AL_n are all less than AL; but as n increases, the K_n and L_n become closer to K and L respectively, as shown above. We may thus choose n sufficiently large that the segments KK_n and L_nL are together less than KL (Fig. 3); that is, $AK < AK_n < AL_n < AL$. Thus, $AK_n : AD < AL_n : AD$. But the latter equals $C_n : P_n$, while $C_n : P_n < AK_n : AD$ (from above). This contradiction makes impossible the assumption that $AK < AL$. Similarly, an argument based on the circumscribed figures shows that $AK > AL$ leads to a contradiction. Thus, $K = L$, as claimed.
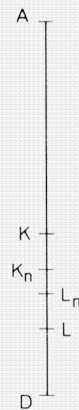
Figure 3

This reconstruction, utilizing procedures well established in QP, CS and $PE I$ and II, arises as the direct formalization of the argument in M, 5, and thus seems likely to represent the approach adopted in the *Equilibria*. In this proof certain facts about centers of gravity have been assumed; for instance, that the center of gravity of a cylinder lies on its axis. Thus, additional theorems like this must have appeared in order to justify their use in the theorem on the paraboloid. In fact, we may identify more specifically which theorems those were. For under the heading "preliminary assumptions" before M, 1 Archimedes lists several theorems on centers of gravity. Some of these receive formal proof in other works: the center of gravity of any figure formed by the removal from a given figure of known magnitude and center of gravity of a part, also of known magnitude and center of gravity (*PE I*, 8); the center of gravity of the triangle (*PE I*, 14) and that of the parallelogram (*ibid.*, 10). Again, a lemma on the proportion of four ensembles of magnitudes whose elements satisfy a specified condition is assumed in the *Method*, but proved as CS, prop. 1. Thus, these are not, properly speaking, "assumptions" – that is, definitions or results suitably basic or obvious to be admitted without proof – like those in *PE I* and *Sphere and Cylinder I*. Rather, they are results whose proofs had already been given in earlier works and could thus be assumed in the *Method*.

The same should be true of the other "assumptions" listed in the *Method*: namely, the centers of gravity of the cylinder, the prism and the cone. These too must have received formal treatment earlier. Although we have no explicit information as to where these appeared, the *Equilibria* would be the appropriate place.[7] As noted, the formal treatment of the paraboloid requires the result for the cylinder, so that this also must have appeared in the *Equilibria,* either as a theorem with demonstration or as a lemma drawn from still another work. The latter alternative, while possible, stretches unnecessarily the hypothesizing of lost treatises, in this case without even indirect evidence for support.

An examination of the formal treatment of these results, worked out in accordance with Archimedean techniques, indicates how they might fit into a sequence

leading to the determination of the center of gravity of the cone. No heuristic version of this result appears in the *Method*, although one might be provided.[8] Indeed, the fact that it is stated as "preliminary" suggests that Archimedes did not handle it by means of the mechanical method of indivisibles.[9] As before, one finds that suitable models for the demonstration may be derived from the proofs in QP and $PE I$ and II. (See Appendix for details.)

As with our reconstruction of the theorem on the paraboloid, each step of the proof of the center of gravity of the cone has been patterned after related formal proofs in the extant mechanical writings. But the treatment of the cone differs from the other in that it requires inscribed figures only. This restriction to one-sided convergence is in keeping with the technique used in Euclid's *Elements* XII and in Archimedes' *PE I* and *II*, and might ostensibly be taken to indicate that the theorem on the cone originally appeared with those early mechanical writings (cf. note 7).[10] But, as we have seen through our remarks on the term "axis," that theorem did not appear before the *Conoids* and thus most likely was first published in the *Equilibria*. As it happens, there *is* a step in this proof which is difficult to handle in the Euclidean manner: the requirement in (4) that the inscribed figure converges to the pyramid. This result is established in XII, 5, but by a method of inscribing figures different from that needed in the theorem on the center of gravity; hence, a new convergence proof must be given. Archimedes' technique of comparing the circumscribed and inscribed figures – establishing that their difference becomes arbitrarily small, so that each sequence converges to the enclosed figure – handles this problem far more efficiently than any adaptation of the Euclidean procedure could. Beyond this, however, one gains no advantage in applying the Archimedean technique of inscribed and circumscribed figures. Indeed, if one were to follow precisely the pattern of QP, 14–15 and actually work out the positions of the centers of gravity of these bounding figures, difficult computations would arise, such as the summation of successive cubes.[11] Thus, the nature of the theorem here recommends not attempting to make full application of the more advanced techniques of QP and CS, even though these would be available to Archimedes at the time of writing the *Equilibria*.

If one accepts our view that the theorem on the center of gravity of the cone, as well as that on the paraboloid, appeared in the lost *Equilibria*, it follows that several other results were given in that work. For the result on the paraboloid, as Archimedes presents it in the *Method* and as we have reconstructed it above, does not use the center of gravity of the cone. The latter result is required for the centers of gravity of the hemisphere (*Method*, prop. 6), the segment of the sphere (prop. 9) and, presumably, also the segments of the ellipsoid and the hyperboloid (props. 10 and 11,

respectively; Archimedes omits the details in these two instances). Thus, Archimedes' purpose in providing the theorem on the cone in the *Equilibria* must have been to effect the formal determinations of the centers of gravity of these other solids. Moreover, as indicated in *Floating Bodies II*, 2, other proofs in the *Equilibria* generalized beyond the results in the *Method*, by removing the restriction that the plane cutting off the segments be perpendicular to the axis. In sum, the *Equilibria* must have been a substantial work, complementary both in structure and magnitude to the *Conoids and Spheroids*.

In a remark at the end of M, 11 Archimedes speaks of "many other theorems" related to those given. One naturally thinks of the centers of gravity of the truncated conoidal segments. In fact, Archimedes characteristically takes up truncated forms after completing the study of basic figures – for instance, the trapezium (*PE I*, 15) follows from the triangle (*ibid.*, 14) and the truncated parabolic segment (*PE II*, 10) follows from the segment (*ibid.*, 8). However, the volumes of the truncated conoids do not appear in *Conoids and Spheroids*, so that one cannot infer that these figures were taken up in the *Equilibria*. Moreover, the centers of gravity of the truncated forms would follow from the known centers of gravity of the conoids; they would not require further application of the mechanical procedure of the *Method* or of the formal procedure based on it. If Archimedes had such results in mind, his comment in M, 11 would mean that "we might take in many other theorems related to these, but whose proofs differ from those whose method has been illustrated in the above." The text here is obscure, however, so that the precise sense could not be made out by Heiberg.

From the version of *Floating Bodies* available to them, Renaissance scholars knew of the relevance of the problem of centers of gravity of solids for Archimedes' work. But lacking the *Method*, they had no direct knowledge of Archimedes' manner of handling these problems, and were thus left to devise proofs based on the methods in the other mechanical writings. Some, like Maurolico, Commandino and Valerio, held closely to the Archimedean method of indirect proof. Others, like Stevin, sought less cumbersome approaches, ultimately working out a form of limiting method.[12] One by one the problems we have indicated were taken up: the cone, the paraboloid, the other conoids, their truncated forms. Commandino's treatment of the cone, for instance, develops along the same steps we have proposed here. But in their treatment of the paraboloid, we can see a conception notably different from that used by Archimedes. As we should expect, they view the paraboloid as bounded above and below by ensembles of cylinders; this much is clear from the procedures in *Conoids*. Although Archimedes' method in *Quadrature of the Parabola*, 14–15 might have suggested to them the argument we have reconstructed for Archimedes, based on

Method, 5, the commentators miss the cue and follow instead a more straightforward, albeit far more cumbersome, route. By computational procedures they determine the exact position of the centers of gravity of the inscribed and circumscribed figures — namely (with reference to Fig. 2 and 3), $AK_n' : AD = 2n - \frac{1}{2} : 3n$ and $AK_n : AD = 2n + \frac{1}{2} : 3n$ From this, one may produce a formal indirect argument or an informal limiting argument to conclude that $AK : AD = 2 : 3$. By contrast, Archimedes' method of weighing the paraboloid against the cone removes the need actually to compute the center of gravity of any solid here.

Archimedes' treatment of center of gravity differs from these more familiar approaches in other respects besides its dispensing with the notion of limit. For one, he never articulates the concept of "moment," the product of magnitude times distance. The center of gravity is always determined through the condition of equilibrium, as proved in *PE I*, 6–7, namely, that when two magnitudes have a ratio reciprocal to that of their respective distances from a certain point, that point is their center of gravity. Likewise, the earlier commentators in the Renaissance start off from this Archimedean technique. Only later, as with Cavalieri, is the center of gravity determined through the summation of the moments of its parts.[13] Thus, in his study of the centers of gravity of solids, the summations done by Archimedes are always of volumes. In the *Method* he avoids having actually to work out the summation by utilizing the geometric properties of suitably related solids; in the *Equilibria*, however, the summations are required. Even for that work, the operations performed are only in the most narrow sense an equivalent to the evaluation of definite integrals, standard in modern expositions of this subject. It is a major distortion of Archimedes' conception to interpret it in such a manner.[14]

Knowing the existence and nature of the *Equilibria*, we may better assess the relation of mechanics to geometry in Archimedes' work. Archimedes engaged in mechanical inquiries with two ends in view: as a route to the discovery of new theorems in geometry; and as a context for the elaboration of formal geometric principles. The *Method* and the *Quadrature of the Parabola* provide examples of the former; the *Equilibria* joins the books of *Plane Equilibria* and *Floating Bodies* to illustrate the latter. But however prominent the mechanical aspect of Archimedes' work may seem, we should not mistake the essentially geometric purpose of even these parts of the corpus. On a comparatively slight foundation of mechanical ideas — in effect, nothing more than the principle of equilibrium — Archimedes built a geometric structure which, even through the ruins which survived antiquity, inspired and challenged geometers centuries afterward.

APPENDIX

We shall outline the proof for the cone in the following steps:

(1) The center of gravity of any prism lies on the plane parallel to its opposite faces which bisects the prism. — The proof is strictly analogous to that for the case of the parallelogram (*PE I*, 9). It is clear that the same result holds for cylinders and cylindroidal solids whose opposite faces are parallel.

Corollary: the center of gravity of a parallelepiped lies at the midpoint of the line connecting the centers of gravity of opposing faces. — This follows from (1), upon viewing the solid as a prism in its three different ways, and from *PE I*, 10 (the center of gravity of the parallelogram).
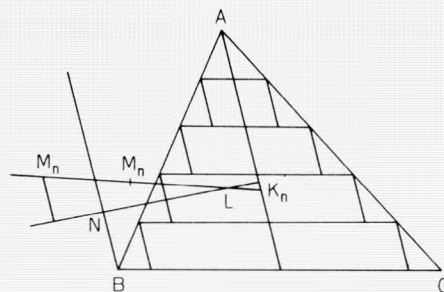
Figure 4

(2) The center of gravity of a prism with triangular faces is the bisector of the line connecting the centers of gravity of the faces. — To prove this, we need only adapt Archimedes' treatment of the center of gravity of the triangle (*PE I*, 13). By (1) the center of gravity of the prism lies in the triangular section formed by the plane bisecting the prism. Let K be the center of gravity of this triangle and K_n the center of gravity of the figure formed by the ensemble of n-many inscribed parallelograms whose sides are parallel to the base or the median of the triangle (Fig. 4); both K and K_n lie on the median. This figure may be used to define a solid comprised of parallelepipeds inscribed in the prism; by the corollary to (1) its center of gravity is K_n. We claim that K is the center of gravity of the prism. As the proof is characteristic of an important aspect of Archimedes' formal method and will come up again later, we shall present it in some detail here. Suppose that the center of gravity of the prism is at L and does not lie on the median. Draw through L the line perpendicular to the median and extend this to meet at N the line parallel to the median through the vertex B of the base. The center of gravity of the difference between the prism and the inscribed solid will be a point M_n on $K_n L$ extended beyond L. As n increases, this difference decreases to become arbitrarily small, so that a point M_n may be found whose projection on LN lies beyond N. This is impossible, however, since all the elements of volume comprising the solid whose center of gravity is M_n lie on the inward side of BN. Hence, L must lie on the median. It follows that $L = K$, as claimed.

(3) The center of gravity of any prism is the bisector of the line joining the centers of gravity of the opposing faces. — This is one of the lemmas stated in the *Method*. Its proof may be given as an adaptation of (2), in which the polygonal

faces are approximated by ensembles of inscribed parallelograms whose centers of gravity become arbitrarily near to that of the face. Similarly, it follows that the center of gravity of a cylinder or of any cylindroidal solid with parallel faces is the bisector of its axis.[15]

In the context of the prism Archimedes introduces a change in terminology, calling the line which joins the centers of gravity of the parallel bounding faces of the prism its "axis" (1, vol. II, p. 432, 433n). Earlier writings, such as Euclid's *Elements* XII, the works on spherics and Archimedes' *Sphere and Cylinder*, assigned this term to the fixed line about which a plane figure rotated to generate a solid of revolution. By contrast, in the tradition of conics one spoke of the "diameter," defined in terms of the sectioning of the cone by a plane. In particular, for the segment of a parabola, the reference line was that drawn from the vertex to the base "parallel to the diameter;" one then proved that this line bisected the base and all chords parallel to the base (*QP*, 3). Archimedes adopts this terminology in *PE II* and *QP*. Later, in the *Conoids*, he uses the same procedure for segments of paraboloids; but there he uses the term "axis" instead of "diameter." This sense of the term "axis" is different from that used in the investigation of the centers of gravity of prisms and cylinders. Had Archimedes proved these theorems *before* the treatise on conoids, he should have defined the "axis" of the segment of the paraboloid as the line between its vertex and the center of the circle or ellipse which forms its base; that this line is parallel to the principal axis of the conoid could then be proved as a theorem. Of course, in the *Equilibria, Floating Bodies* and the *Method* the term "axis" had to be understood in relation to the center of gravity. Thus, the theorems on the prism and cylinder and cone, which would occasion this redefinition of the term, appear not yet to have been published at the time of *Conoids*. This confirms our judgment (see note 7) assigning those theorems to the *Equilibria*.
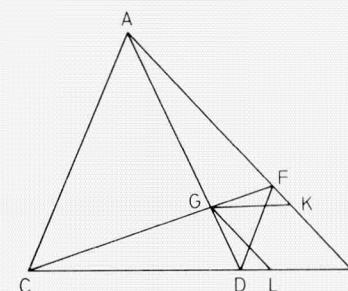

Figure 5

(4) The center of gravity of any pyramid having a polygonal base lies on the line joining the vertex with the center of gravity of the base. — We inscribe in the pyramid an ensemble of prismatic solids whose lateral faces are parallel to the medial line and whose opposite faces are parallel to the base. By (3) the centers of gravity of these prisms, and hence that of their sum, fall on the medial line. An indirect argument, like that in (2), establishes our claim. The same result holds for the case of the cone and for any other pyramidal solid with curvilinear base.

Corollary: the center of gravity of a pyramid with triangular base divides the medial line in the ratio 3:1 (where the greater segment lies in the direction of the vertex). — In Fig. 5 let AB and CB be medians of two faces of the pyramid, F and D be the centers of

gravity of these same faces, respectively. The triangles ADB and CFB lie in the same plane, so that their sides AD and CF will intersect; denote this point G. By (4) G is the center of gravity of the pyramid; we must show that $AG : GD = 3 : 1$. As F and D are the centers of gravity of triangles, $AF : FB = CD : DB = 2 : 1$ (cf. *PE I*, 15). We draw GK and GL parallel to CB and AB, respectively. By similar triangles, $FK : FB = GK : CB$ and $DL : DB = GL : AB$. Since $GK = DB - DL$ and $GL = FB - FK$, one will obtain that $FK : FB = DL : DB = 1 : 4$. Since $AG : GD = AK : KB = AF + FK : BF - FK$, one obtains that $AG : GD = 3 : 1$, as claimed. Although manipulations of this type are unwieldy under the Greek mode of representing magni- tudes, they are nevertheless well within the compass of Archimedes' technique, as the arguments on the centers of gravity of the trapezium and the truncated parabolic segment make evident (*PE I*, 15 and *II*, 10, respectively). We may obtain this result from the same construction in a different manner by following the pattern of the determination of the center of gravity of the triangle, as preserved by Pappus (*Collection* VIII, ed. F. Hultsch, p. 1036; cf. p. 1073n). Since triangles ABC and FBD are similar, $AC : FD = AB : FB = 3 : 1$. Since triangles ACG and FDG are similar, $AC : FD = AG : GD$. Hence, $AG : GD = 3 : 1$, as claimed.
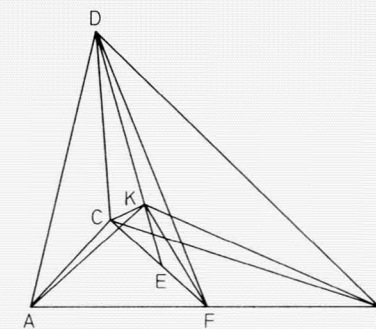

Figure 6

This step of the argument may be effected alternatively, along the lines presented by Commandino. Assume the pyramid $ABCD$ and set E as the center of gravity of the face ABC (Fig. 6); having already proved that the center of gravity of the pyramid, point K, is the intersection of the lines drawn from each vertex to the center of gravity of the opposite face (6, prop. xvii; cf. our step 4), Commandino now proposes that K divides each such line in the ratio 3 : 1 (6, prop. xxii). Using K as vertex he subdivides pyramid $ABCD$ into four pyramids, $KABC, KBCD, KABD, KACD$. He next establishes that these four pyramids are equal. For if the median line CEF is drawn, to bisect edge AB at F, triangles AFC and BFC will be equal; thus pyramids $DAFC$ and $DBFC$ are equal, and pyramids $KAFC$ and $KBFC$ are equal. As triangles DAF and DBF are equal, so pyramids $KDAF$ and $KDBF$ are equal. Now, pyramid $DAFC$ is the sum of $KAFC, KDAF$ and $KACD$; similarly, $DBFC$ is the sum of $KBFC, KDBF$ and $KBCD$. It follows that $KACD = KBCD$, after subtraction. Proceeding in the same manner, he can obtain that the other two pyramids, $KABC$ and $KABD$, are also equal to these. Since, further, $ABCD : KABC = DE : KE$, he can conclude that $DK : KE = 3 : 1$. — Of course, we cannot know which approach, if either, Archimedes employed. Presumably, his determination of the center of gravity of the triangle might, by analogy, serve as guide. But, unfortunately, Archimedes proves only that its center of gravity lies on the median line (*PE I*, 13) and is thus the point of intersection of the three median lines (*ibid.*, 14). That this point divides each median in the ratio 2 : 1 is assumed in *PE I*, 15 (as also in *QP*, 6 and *M*, 1); but the proof is missing from the extant versions and Eutocius does not provide one in his commentaries. If we assume that Archimedes knew only

one of these three variations, to the exclusion of the others – an assumption not necessarily correct – the second, modeled after the related theorem in Pappus, appears to have the strongest textual support. At any rate, this step is the only discrepancy between Commandino's proof of the center of gravity of the cone and the present one (produced without prior knowledge of his). This may be taken to indicate that the overall treatment correctly reproduced Archimedes' method.

Figure 7

(5) The center of gravity of a pyramid with polygonal base divides the line joining the vertex with the center of gravity of the base in the ratio 3 : 1. – Given the pyramid of vertex A and base BCD. . ., we denote by L the center of gravity of the base and by means of the lines $LA, LB, LC, LD, \ldots$ we subdivide the pyramid into pyramids with triangular bases (Fig. 7). If L_1 is the center of gravity of any of these triangular bases, the center of gravity of the corresponding pyramid K_1 divides AL_1 such that $AK_1 : K_1L_1 = 3 : 1$ (by the previous corollary). If any two such points L_1, L_2 are taken, it follows that the line segment determined by the corresponding points K_1, K_2 is parallel to L_1L_2. Thus, all the centers of gravity $K_1, K_2, K_3, \ldots$ lie in the same plane. The center of gravity K of the entire pyramid must lie in this same plane; by (4) it lies also on the line AL. By similar triangles, $AK : KL = AK_1 : K_1L_1 = 3 : 1$, as claimed.

Figure 8

(6) The center of gravity of the cone divides its axis in the ratio 3 : 1. – The proof employs an indirect argument close in style to that in step (2). Let L be the center of the base, A the vertex. If a regular polygon is inscribed in the base and the pyramid with this polygon as base is completed, the center of gravity of the pyramid will be the point K which divides AL in the ratio 3 : 1 (by the previous result). We claim the center of gravity of the cone is K. If not, let it be at a different point M

(Fig. 8). We draw AM extended to meet the base at N; LN is extended to meet the circle at B; and KM is drawn extended to meet at P the line through B parallel to AL. Now, K is the center of gravity of any pyramid whose base is the regular polygon of n-many sides centered on L. Thus, the figure which is the difference between the cone and this pyramid will have its center of gravity at P_n, a point on the line segment from M in the direction of P. By increasing the number of sides n, the volume of this difference becomes arbitrarily small. Setting X the volume of the cone and Y_n the volume of the pyramid, we define the volume Z by the proportion $KM : MP = Z : Y_k$ (where k has an arbitrary fixed value, say 4). We may find n such that X - Y_n is less than Z (and $n > k$). Then the center of gravity of $X - Y_n$ will lie at P_n such that $KM : MP_n = X - Y_n : Y_n < Z : Y_k = KM : MP$. It follows that $MP_n > MP$. But this is impossible, since all the elements which comprise $X - Y_n$ lie on the inward side of P. This contradiction establishes that K lies on the axis AL; similar indirect arguments show that $K = L$, as claimed. This completes our outline of the proof of the center of gravity of the cone.

NOTES

1. On these commentators, see Wieleitner (13), Boyer (3, pp. 98–106), Clagett (5, vol. III) and Rose (12, ch. 8, 9, esp. pp. 167–8, 200–202).
2. The description of this manuscript is provided by Heiberg in *Archimedes* (1, vol. III, pp. lxxxv-xc), by Heath (9, vol. II, pp. 25–27; 10, *Supplement*, pp. 5–6) and by Dijksterhuis (7, ch. II).
3. Archimedes (1, vol. II, p. 350). One should beware using the editions of Heath in any close study of *Floating Bodies* or the *Method*. For his edition appeared before Heiberg's second edition of Archimedes, in which the Greek text of the former and a revised text of the latter were given. The *Method* was unknown to the Renaissance scholars, while *Floating Bodies* existed only in the Latin translation made by William of Moerbeke in 1269; this version lacks the portion of prop. 2 where the lost *Equilibria* is cited (Clagett, 5, II, p. 365).
4. Child (4, p. 513) provides an alternative treatment in which the paraboloid in position balances a triangular prism in position; the prism is such that each of its rectangular sections equals the circular section of the paraboloid equidistant from the fulcrum. His explanation for Archimedes' not using this allegedly simpler method, i.e., that it would require a drawing in perspective, is incredible. Rather, Archimedes was reluctant to introduce the assumption of a rectilinear figure equal to a circle – indeed, this is inferable from a remark in the preface to the *Method*. On the other hand, had he possessed alternative methods for the theorems in the *Method*, his choice of one form rather than another might well have been arbitrary.
4a. Archimedes retains the notion of the balance in the area-theorem of QP, 14–16. But I maintain that this was considered improper according to the formal standards of his time (11, sect. III. 1). In no other work, save the *Method*, does he use mechanical assumptions in the proof of a purely geometric problem.
5. Archimedes' proof that the center of gravity of the triangle lies on its median (*PE I*, 13) employs an analogous argument.

6. More specifically, the sequence of points K_n' converges monotonically to K from the direction of A, while the sequence K_n converges monotonically to K from the direction of D.
7. Heiberg (1, vol. II, p. 548) makes such a conjecture, but without supporting argument. By contrast, Arendt wishes to limit the content of the *Equilibria* to just those theorems on centers of gravity presented in the *Method* (2, pp. 299–300); he refers the theorems on the cylinder, prism and cone to another lost work, the *Mechanica*, of which he takes *PE II* to be a fragment. But, as I argue in a forthcoming paper on the chronological ordering of Archimedes' works (11, sect. II. 4), *PE II* is best associated with the more elementary techniques of the early works in the corpus. The theorems at issue here, whose proofs we sketch below, appear to conform in content and technique to the proofs in the *Equilibria*, without doubt a late work in the corpus.
8. For instance, one might balance the cone in position by a prism in position, whose parallel faces are a funnel-shaped plane figure defined by the space between two equal semi-parabolas. The center of gravity of the semi-parabola is readily obtained from that of the parabolic segment (*PE II*, 8), from which that of the prism follows. Of course, it would seem improper to use this more complex figure in the study of the problem of the simpler figure of the cone.
9. Child provides a mechanical argument to derive the volume of the cone from the moment of the triangle (4, p. 512). His method is awkward, however, and utilizes concepts of moment and of equilibrium between non-homogeneous magnitudes which Archimedes certainly would not have allowed. Still, the argument may be remedied to produce a valid treatment involving the equilibrium of the cone suspended at a fixed distance with a triangular prism in position. Archimedes' reason for not producing such a treatment is, of course, that the volume of the cone was by his time a long-familiar result. The theorems in the *Method* are all original discoveries by Archimedes. We cannot know for certain, although it is probable, that Archimedes recognized the possibility of such an alternative treatment of the cone.
10. In "Archimedes and the *Elements*" (11) I argue that a criterion of precisely this kind can help distinguish early from late parts of the Archimedean corpus.
11. Galileo adopts this approach in his derivation of the center of gravity of the cone (8, Appendix: ". . . demonstrations concerning the center of gravity of solids," prop. 6–8). The summation of consecutive cubes was known in antiquity (Heath, 9, vol. I, pp. 108–110), so that a similar proof may have been possible for Archimedes.
12. The treatments by Maurolico and Commandino are discussed in some detail by Wieleitner (13) and are presented, along with those by Stevin and Valerio, by M. Baron, *Origins of the Infinitesimal Calculus*, Oxford, 1969, ch. 3. A brief account appears in Boyer (3, pp. 98–106).
13. For instance, given two magnitudes the moments of whose corresponding indivisible parts have a specified relation, Cavalieri concludes that the moments of the wholes are similarly related; this is the basis of his

derivation of the "Guldin rule" (*Geometria degli Indivisibili*, ed. L. Lombardi-Radice, Turin, 1966, Appendix: "Polemic with Guldin," pp. 847ff.).
14. This distortion mars the discussion of the *Method* by Child (4, pp. 509–520) and Zeuthen (summarized by Heath, 10, *Supplement*, pp. 7–10). Under the definition by moments, one determines the center of gravity of the paraboloid by computing $\int x^2 dx / \int x\, dx$. Archimedes solves this problem in the *Method* through the ratio of the volumes of the cone and the paraboloid. Formally, perhaps, the two approaches may be equivalent. But the essence of Archimedes' conception is lost under such an interpretation.
15. It is odd that Archimedes states the result for the cylinder *before* that for the prism, in the lemmas to the *Method*, even though it would appear that the former result is derivable as a corollary to the latter, as we have presented them. Actually, Archimedes' ordering of the lemmas need not be significant. For instance, he lists the center of gravity of the triangle before that of the parallelogram, even though his own demonstrations of them involve using the latter (*PE I*, 10) in the proof of the former (*PE I*, 13).

BIBLIOGRAPHY

1. Archimedes, *Opera Omnia*, ed. J. L. Heiberg, 3 vol., Leipzig, 1910–15 (repr., Stuttgart, 1972).
2. Arendt, F., "Zu Archimedes," *Bibliotheca Mathematica*, 1913–14, 14³, pp. 289–311.
3. Boyer, Carl, *Concepts of the Calculus*, 2. ed., New York, 1949 (repr. 1959).
4. Child, J. M., "Archimedes' Principle of the Balance, and Some Criticisms upon it," in Charles Singer, ed., *Studies in the History and Method of Science*, Oxford, 1921, vol. 2, pp. 490–520.
5. Clagett, Marshall, *Archimedes in the Middle Ages*, vol. I, Madison, 1964; vol. II, Philadelphia, 1976; vol. III, Philadelphia (in press).
6. Commandino, F., *Liber de centro gravitatis solidorum*, Bologna, 1565.
7. Dijksterhuis, E. J., *Archimedes*, New York, 1957.
8. Galileo, *Two New Sciences*, ed. S. Drake, Madison, 1974.
9. Heath, T. L., *A History of Greek Mathematics*, 2 vol., Oxford, 1921.
10. Heath, T. L., *The Works of Archimedes*, Cambridge, 1897 (with the *Supplement, The Method of Archimedes*, 1912), (repr., New York, n.d.).
11. Knorr, Wilbur, "Archimedes and the *Elements*: Proposal for a Revised Chronological Ordering of the Archimedean Corpus," *Archive for the History of Exact Sciences* (in press).
12. Rose, Paul L., *The Italian Renaissance of Mathematics*, Geneva, 1975.
13. Wieleitner, H., "Das Fortleben der archimedischen Infinitesimal-methoden . . .," *Quellen und Studien*, 1931, 1:*B*, pp. 201–220.

Brooklyn College
City University of New York
Brooklyn, New York, USA

The Old Intelligencer

I liked it as it was. I could read it in a few minutes, laugh at the jokes & throw it away. Now it is just another journal, of which we already have too many. What you had before was a good advertisement for Springer, I doubt if the new one is. I have sent an order form to your NY office in case it turns out better than I expect.

Sir Edward Bullard
Scripps Institution of Oceanography
La Jolla, California
October 6, 1977

****

From a letter of Frobenius to Dedekind
May 7, 1896

. . . According to the latest Lampe-Wanger Jahrbuch, Burnside has apparently published the theorem that for any Abelian group

$$x_{AB^{-1}} = \sum_R \chi(R) x_R.$$

It was in "On a property of certain determinants" [Messenger of Mathematics, 1893, XXIII, p. 112]. I hope this doesn't upset you too much. This is the same Mr. Burnside who, for some years now, has been annoying me by rediscovering every every single theorem I have published in group theory. What's more, he doesn't just rediscover them, but rediscovers them without exception in the very same order in which they have been published: first my proof of Sylow's theorem, then the theorem on groups of quadratic order, then the one on groups of order $p^{\alpha}q$, then the one on groups whose order is a product of 4 or 5 primes, etc., etc. Anyway, it is one of those remarkable and wonderful examples of psychic harmony (*Seelenharmonie*) which can only occur in England or possibly in America.

Puzzles

R. M. Smullyan

Raymond M. Smullyan of The City University of New York has given us the following puzzles. The first two are from his book What Is The Name Of This Book? to be published in May by Prentice-Hall, Inc. The last two are, in Professor Smullyan's words "from my manuscript 'To Mock A Mocking Bird' now in preparation. The idea of the book is to develop some mathematics and logic out of recreational puzzles. So, the book starts out as a pure puzzle book, and gradually spreads out to various areas, including such topics as infinite sets, mathematical induction, some new results in double induction, Konig's infinite lemma, fixed point theorems, Boolean Algebras, First-Order logic (with proofs of completeness and compactness), some theory of formal systems with at least one proof of an undecidability result, and (if space permits) a chapter on the Axiom of Choice with a proof of at least one maximal principle!"

1. *Knights, Knaves and Normals*
   A certain island is inhabited by *knights* who always tell the truth, *knaves* who always lie and *normals* who sometimes lie and sometimes tell the truth. You yourself are an inhabitant of the island and have fallen in love with the King's daughter and wish to marry her. However, the King will allow his daughter to marry only a "normal" (knights, he feels, are too sanctimonious, and knaves are too treacherous). Therefore, if you can convince the King that you are normal, then you may marry his daughter; otherwise, not. Suppose that you in fact are normal, and you are allowed to make as many statements before the King as you wish. Obviously two statements are enough to convince the King that you are normal—just make one statement which is obviously true and one statement which is obviously false. Suppose however:
   (a) You are not allowed to make any false statements before the King (if the King catches you in a falsehood, you get executed on the spot!). Is it now possible to convince the King that you are normal? If so, what is the smallest number of true statements you can make which will convince the King that you are normal?
   (b) Suppose, instead, the King (for some perverse reason) allows you to make only false statements. Is it now possible to convince the King that you are normal? If so, in how many statements?
   (c) Suppose, instead, the King requires proof not only that you are normal, but also that you are intelligent. He therefore demands that you make only one statement, and that it satisfies the following two conditions:
   I   From it, the King can deduce that you must be normal,
   II   From it, the King cannot deduce whether the statement is true or false.
   The problem is to construct a statement satisfying these two requirements.
   (d) Suppose, instead, the King wants his daughter to marry only someone who is *not* normal. You are allowed to make as many statements as you like, and there is no restriction on their truth-values. Assuming you are, in fact, not a normal, what is the smallest number of statements necessary to convince the King that you are not normal?

2. *The Island of Zombies*
   Another island is inhabited exclusively by *humans*, who always tell the truth, and *zombies*, who always lie. Although the natives all understand English perfectly, there is an ancient taboo on the island forbidding them to use English words in their speech. However, whenever you ask them a question answerable by *yes* or *no*, they will answer with either "Bal" or "Da", one of which means *yes* and the other *no*, but we do not know which is which. There is a rumor that there is gold buried on this island. All the natives know whether the rumor is true or false.
   How, in only one yes-no question (which will be answered *Bal* or *Da*) can you find out if there is gold on the island?

3. *A Gödelian Puzzle*
   There is another marvelous island — Island G — inhabited exclusively by knights (who always tell the truth) and knaves (who always lie). (There are no normals on this island). Certain knights who have been *proved* to be knights are called *established* knights, and certain knaves who have been *proved* to be knaves are called *established* knaves. For the moment, the question is open whether all the knights are established and whether all the knaves are established.
   The inhabitants of this island have no names, hence communication has always been somewhat of a problem. One day, however, a Gödely being appeared from the sky and assigned to every inhabitant a number (positive integer) henceforth known as the *Gödel number* of the person. For any Gödel number n, we let P_n be the person who has this Gödel number.
   The registrar of this island keeps a certain book called the *book of sets.*. The pages are numbered consecutively, and on each page is written a description of a set of positive integers. The sets which are described on the various pages are called the *listed* sets. We call n an *index number* if n is the number of some page of the book, and we let S_n be the set listed on page n. If n happens to be a member of the set S_n, then we call n an *extraordinary* number. For any index number n, a number m is called an *associate* of n if m is the Gödel number of a person who makes the claim that n is an extraordinary number (of course the claim could be false; P_m might be a knave). We are given the following five conditions:

C_1: The set of Gödel numbers of all the established knights is a listed set.
C_2: The set of Gödel numbers of all the established knaves is a listed set.
C_3: The complement of any listed set is a listed set.
C_4: Every index number h has at least one associate.
C_5: For any listed set A, the set of all numbers which have at least one associate in A is again a listed set.
   (a) (After Gödel) Prove that at least one of the knights is unestablished and that at least one of the knaves is unestablished.
   (b) (After Tarski) Can it be determined whether or not the set of Gödel numbers of all the knights on the island is a listed set?

4. *To Mock a Mocking-Bird*
   We are given a denumerable sequence B_1, B_2, B_3, . . . B_n, . . . of birds. The sequence contains no repetitions. Whenever you call out a number (positive integer) to a bird B, she responds by calling some number back to you. If she calls back the same number, then we say that the bird is *fond* of that number — or *fixated* on that number. If a bird B_n is fond of her own index n then we say that B_n is *egocentric*.
   We shall use functional notation: for any bird B and number n, we let B(n) be B's response to n. Thus B is *fond* of n iff B(n) = n, and B_n is *egocentric* iff $B_n(n) = n$. We are given the following two conditions:

C_1: One of the birds M is called a *mocking-bird*, because for every number n, M's response to n is the same as B_n's response to n, $(M(n) = B_n(n))$.
C_2: The functions indexed by the birds are closed under composition — i.e. for any birds A and B, there is a bird C such that for every n, A(B(n)) = C(n).
   The problem is to prove that every bird is fond of at least one number, and that at least one bird is egocentric.

5. *The Magic Garden of George B.*
   A little boy George B. has a garden of magic flowers. Each flower can change color from day to day, but it assumes only one of the colors blue or red. On any day, a flower is either blue the entire day or red the entire day. We are given the following 3 conditions:

C_1: No two distinct flowers have the same color on all days — stated otherwise, given any two distinct flowers A and B, there is at least one day where one of them is red and the other is blue.
C_2: For any flowers A,B there is a flower C such that C is blue on all and only those days when A,B are both red.
C_3: There are somewhere between 200 and 300 flowers in the garden.
How many flowers are there in the garden?

City University of New York
New York, USA

THE MATHEMATICIAN'S ART OF WORK

BY J. E. LITTLEWOOD

I WILL BEGIN by saying, with a double motive, that there are a lot of queer people in the world. I remember a report of a man who was three times saved from drowning in a day, bathed once more and *was* drowned. The same year there was the case of a man, whom I think of as having spent his life in the British Museum, who conceived the idea of a seaside holiday, had himself rowed out, dived overboard, and being unable to swim, was drowned.

A former pupil began brilliantly; he took a pure research post after his Ph.D. under me, and had 6 years of research. The work became dull though copious, and finally ended, and when I then met him he was on the point of a nervous breakdown. I then discovered that he had worked continuously for the 6 years for 365 and a-quarter days a year. If he had done the work in the reverse order he would have been a Fellow of the Royal Society.

Shortly before World War I the psychologist, Boris Sidis, subjected his son to a theory of education. By the age of 19 the boy had become outstanding in a number of subjects. According to the theory no strain was involved, and in fact the boy did not obviously overwork and was also good at athletics. Some time after the war, when he was about 30, he was met by visitors from England. He held an ill-paid post with unexacting duties, refused to be promoted, and said that his object in life was never to have to *think* again.

These are awful examples. On the other hand, there are successfully creative people with strange methods of work. I know of a man who works only two days a week, of one who can work only in a cabaret, of one who has a wine-bottle by his desk. The economist, Marshall, though he had been through the mill of the Cambridge Mathematical Tripos and was second Wrangler, could not, or at any rate did not, work in later life for more than fifteen minutes at a stretch.

There are two morals to all this. If a young man feels he is not at home in the world, or that his instincts of how to work are abnormal, there is no reason for him to worry unduly. On the other hand he would be wise to find out what the usual methods are and give them a prolonged trial (less than a month is no good at all) before finally committing himself. There can be powerful illusions on such points, which I will come to later.

Creativity

At the lowest level there is an element of creativity in much ordinary conversation. We do not think what we are going to say and then say it, the experience is subjectively simple, and what is said emerges from the subconscious into the conscious. Long experience has established a working liaison between the two, but if it fails, one becomes "tongue-tied."

At the other extreme of creativity, there are the "great" creations, of something totally new and unexpected, and also of great importance, and seminal. We should all feel that the difference is one of kind and not of degree. (If I may frivolously digress, do you know the question: is the difference between a difference of degree and a difference of kind a difference of degree or a difference of kind? The *answer*, of course, is elementary.)

Much lies between the extremes. In any new form of mental activity, however man-made, the niche fills with people whose capacity is many orders of magnitude above the average. It has been said that we use only a small part of our brains; these facts are perhaps evidence for the idea. Let us run over some cases.

Oscar Wilde could cast his eyes down the pages of a novel, and in five minutes pass an examination on the contents.

The fantastic performances of musical prodigies and calculating boys are well-known. They have an intense interest in their game to the exclusion of all else, but the facts would be incredible if they did not happen. The calculators split into above average intelligence and below. The former lose interest when they realize that anyone can get his results

slowly by routine methods; they reach their peak at about the age of 4. Gauss was a case of this, though he did do a good deal of numerical calculation throughout his life: possibly he found in it a relaxation like that provided today by crosswords; but Gauss was a law to himself. A contemporary of Gauss, Däse, was of low intelligence, kept his capacity all his life, and was actually employed by Gauss to make factor-tables. Bidder, himself a highly intelligent case, had a daughter with a rather different faculty, but one out of all relation with ordinary people. She knew the current 707 digits for π, could begin at any point and read them off *either forwards or backwards*. She was studied by philosophers and psychologists in Cambridge, but she was quite unable to explain how she did it.

Computer theory has thrown up a class with suitable gifts, and they are not necessarily very good at mathematics. There is a class, not apparently very distinguished intellectually, which — as a recent experiment showed — can do difficult crossword puzzles with almost complete certainty and in an incredibly short time.

There are people who can learn a new language in a week, but most adults are poor linguists. Children, on the other hand, if suitably exposed, can be fluently trilingual by the age of six (more than three languages creates confusion). I don't know how to place this difference, but clearly this is the right way to be trilingual.

None of all this is highly creative. But, still between the extremes, there is the army of people very gifted but short of genius. Though the importance of their creations falls short of the highest, I think the psychology involved is pretty much the same. A *sine qua non* is an intense conscious curiosity about the subject, with a craving to exercise the mind in it, quite like physical hunger. Love of truth — and all that — may co-exist, but I deny that it is the driving force. (To digress on physical analogies, a "hunch" — an idea for which one can give no reason — seems analogous to smell.) Given the strong drive, it communicates itself in some form to the subconscious, which does all the real work, and would seem to be always on duty. Lacking the drive, one sticks. I have tried to learn mathematics outside my fields of interest; after any interval I had to begin all over again.

Four Phases

It is usual to distinguish four phases in creation: preparation, incubation, illumination, and verification, or working out. For myself I regard the last as within the range of any competent practitioner, given the illumination.

Preparation is largely conscious, and anyhow *directed* by the conscious. The essential problem has to be stripped of accidentals and brought clearly into view; all relevant knowledge surveyed; possible analogues pondered. It should be kept constantly before the mind during intervals of other work. This last is advice from Newton.

Incubation is the work of the subconscious during the waiting time, which may be several years. Illumination, which can happen in a fraction of a second, is the emergence of the creative idea into the conscious. This almost always occurs when the mind is in a state of relaxation, and engaged lightly with ordinary matters. Helmholtz's ideas usually came to him when he was walking in hilly country. There is a lot to be said for walking during rest periods, unpopular as the idea may be. Incidentally the relaxed activity of shaving can be a fruitful source of minor ideas; I used to postpone it, when possible, till after a period of work. Illumination implies some mysterious rapport between the subconscious and the conscious, otherwise emergence could not happen. What rings the bell at the right moment?

I recently had an odd and vivid experience. I had been struggling for two months to prove a result I was pretty sure was true. When I was walking up a Swiss mountain, fully occupied by the effort, a very odd device emerged — so odd that, though it worked, I could not grasp the resulting proof as a whole. But not only so; I had a sense that my subconscious was saying, "Are you *never* going to do it, confound you; try this."

MY DESCRIPTION so far has been appropriate to science or mathematics, one single idea. In the case of a symphony, incubation would be a more continuous process and many separate illuminations would be called for. And then the final miracle of a great symphony is the welding into an organic whole.

Beethoven's notebooks show that — with some remote objective vaguely in view — he would start with
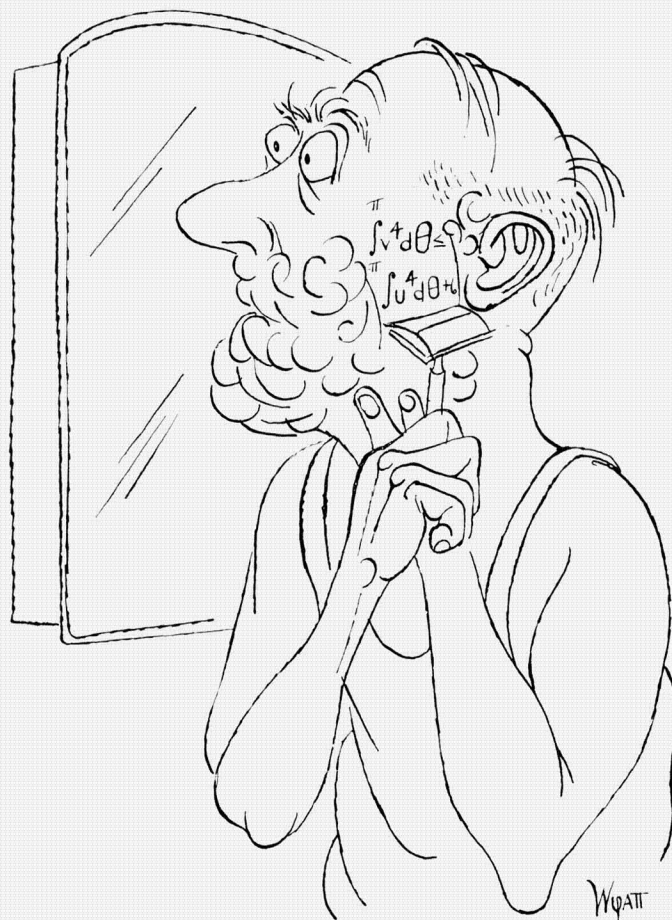
... the relaxed activity of shaving can be a fruitful source of minor ideas; I used to postpone it, when possible, till after a period of work.

deliberate crudities, and approach the final work through blunders and repeated alterations. A recent surprise has been that the apparently completely spontaneous Dvořák was a quite similar case.

Some Anecdotes

KEKULÉ's Benzene ring came in a dream. It is proverbial that sleep alters thoughts and decisions, but I believe high creation in dreams is very rare. William James used to have what seemed vitally important ideas in dreams, but always forgot them on waking. He decided to write down such dreams, and succeeded in doing so on the next occasion. In the morning he read: "Higamus hogamus, woman is monogamous; Hogamus higamus, man is polyga-

mous." It is not too bad: it has both form and content.

MENDELEEV was a conscientious Professor of Chemistry with the urge to do the utmost for his pupils. He collected together likenesses of elements and tried to number them helpfully so as to be mnemonics. The final upshot was the Periodic Law.

LOBACHEVSKI had to teach Geometry; and began by an intensive critical study of Euclid, acting as devil's advocate. This approach to the anomalous parallel axiom resulted in his non-Euclidean Geometry.

M. RIESZ had conceived a beautiful collection of theorems. He could prove them all if he could only prove a very special and innocent-looking case of one of them. Specifically: if $f(z) = u + iv$ is regular in $|z| \leq 1$ and $f(0) = 0$ then we are to have

$$\int_{-\pi}^{\pi} |v|^4 \, d\theta \ \leq A \int_{-\pi}^{\pi} |u|^4 \, d\theta.$$

One day, having to give an examination, he was playing about with $\int_{-\pi}^{\pi} f^4 \, d\theta$ instead of the familiar $\int_{-\pi}^{\pi} |f|^4 \, d\theta$.

This gives $\int_{-\pi}^{\pi} v^4 \, d\theta = -\int_{-\pi}^{\pi} u^4 \, d\theta + 6\int_{-\pi}^{\pi} u^2 v^2 \, d\theta$.

Using Cauchy's inequality on the least provocation was second nature to him. Doing this and changing the *minus* into a *plus* to fit higher indices, he had

$$\int_{-\pi}^{\pi} v^4 \, d\theta \leq \int_{-\pi}^{\pi} u^4 \, d\theta + 6 \left(\int_{-\pi}^{\pi} u^4 \, d\theta \int_{-\pi}^{\pi} v^4 \, d\theta \right)^{\frac{1}{2}},$$

and saw his proof staring him in the face.

A possible moral of the last three instances is that teaching activities may pay off in pure research.

In passing, I firmly believe that research should be offset by a certain amount of teaching, if only as a change from the agony of research. The trouble, however, I freely admit, is that in practice you get either no teaching, or else far too much.

ERASMUS DARWIN held that every so often you should try a damn-fool experiment. He played the trombone to his tulips. This particular result was,

Erasmus Darwin held that every so often you should try a damn-fool experiment. He played the trombone to his tulips.

in fact, negative. But other incredibly impudent ideas have succeeded. An Italian physicist had rigged up two screens and transmitted electric impulses from one to the other. If he had spoken to one of the screens he would have invented the telephone, but with his sound physical sense he of course did nothing of the kind. The phonograph and telephone are surely most impudent ideas.

There is a superb example, which I have described as the most impudent idea in Mathematics, with very important consequences; it is too technical for this paper, but you will find it discussed on pages 20-23 of my *Mathematician's Miscellany*.

On Being a Mathematician

There is much to be said for being a mathematician. To begin with, he has to be completely honest in his work, not from any superior morality, but because he simply cannot get away with a fake. It has been cruelly said of arts dons, especially in Oxford, that they believe there is a polemical answer to everything; nothing is really *true*, and in controversy the object is to prove your opponent a fool. We escape all this. Further, the arts man is always on duty as a great mind; if he drops a brick, as we say in England, it reverberates down the years. After an honest day's work a mathematician goes off duty.

Mathematics is very hard work, and dons tend to be above the average in health and vigor. Below a certain threshold a man cracks up; but above it, hard mental work *makes* for health and vigor (also — on much historical evidence throughout the ages — for longevity). I have noticed lately that when I am working really hard I wake round about 5:30 a.m., ready and eager to start; if I am slack, I sleep till I am called. I mentioned this to a psychological doctor, who said it was now a known phenomenon.

There is one drawback to a mathematical life. The experimentalist, having spent the day looking for the leak, has had a complete mental rest. A mathematician's normal day contains hours of close concentration, and leaves him jaded by the evening. To appreciate something of high aesthetic quality needs close attention, easy to the unfatigued; but a strain for the fatigued mathematician. (Music seems a happy exception to this.) This is why we tend to relax either on mild nonfiction like biographies, or — to be crude, and to the derision of arts people — on trash. There is, of course, good trash and bad trash.

THE HIGHER MENTAL ACTIVITIES are pretty tough and resilient, but it is a devastating experience if the drive does stop, and a long holiday is the only hope. Some people do lose it in their forties, and can only stop. In England they are a source of Vice-Chancellors.

MINOR DEPRESSIONS will occur, and most of a mathematician's life is spent in frustration, punctuated with rare inspirations. A beginner can't expect quick results; if they are quick they are pretty sure to be poor.

To digress on this point, the ideal line for a supervisor with a really promising man is to give him two subjects: one, ambitious; the other, one that the supervisor can judge to be adequate for a Ph.D. (even if he has to do the thing himself first).

When one has finished a substantial paper there is commonly a mood in which it seems that there is really nothing in it. Do not worry, later on you will be thinking "at least I could do something good *then*." At the end of a particularly long and exacting work there can be a strange melancholy. This, however, is romantic, and mildly pleasant, like some other melancholies.

Research Strategy

With a good deal of diffidence I will try to give some practical advice about research and the strategy it calls for. In the first place research work is of a different order from the "learning" process of pre-research education (essential as that is). The latter can easily be rote-memory, with little associative power: on the other hand, after a month's immersion in research the mind knows its problem much as one's tongue knows the inside of one's mouth. You must also acquire the art of "thinking vaguely," an elusive idea I can't elaborate in short form. After what I have said earlier, it is inevitable that I should stress the importance of giving the subconscious every chance. There should be relaxed periods during the working day, profitably, I say, spent in walking.

HOURS A DAY AND DAYS A WEEK On days free from research, and apart from regular holidays, I recommend four hours a day or at most five, with breaks about every hour (for walks perhaps). If you don't have breaks you unconsciously acquire the habit of slowing down. Preparation of lectures counts more or less as research work for this purpose. On days with teaching duties, I can only say, be careful not to overdo the research. The strain of lecturing, by the way, can be lightened if you apply the golfing maxim: "don't press." It is, of course, hard not to. Don't spend tired periods on proof correction, or work that needs alertness; you make several shots at an emendation that you would do in one when fresh. Even in making a fair copy one is on the *qui vive* for possible changes.

Either work all out or rest completely. It is too easy, when rather tired, to fritter a whole day away with the intention of working but never getting properly down to it. This is pure waste, nothing is done, *and* you have had no rest or relaxation. I said "work all out": speed of associative thought is, I believe, important in creative work; another elusive idea, with which my psychological doctor agrees.

For a week without teaching duties — and here I think I am preaching to the converted — I believe in one afternoon and the following day off. The day off need not necessarily be Sunday, but that has a restful atmosphere of general relaxation, church bells in the distance, other people going to church, and so on. The day, however, should stay the same one of the week; this establishes a rhythm, and you begin relaxing at lunch time the day before.

At one time I used to work 7 days a week (apart, of course, from 3-week chunks of holidays). I experimented during a Long Vacation with a Sunday off, and presently began to notice that ideas had a way of coming on Mondays. I also planned to celebrate the arrival of a decent idea by taking the rest of that day off. And then ideas began coming also on Tuesday.

MORNING VERSUS EVENING Before World War I it was usual in Cambridge to do our main work at night, 9:30 to 2:00 or later. Time goes rapidly — one has a whiskey and soda at 11:30 and another later — and work *seems* to go well and easily. By comparison the morning seems bleak and work a greater effort. I am sure all this is one of the many powerful illusions about creative work. When put out of action by a severe concussion in 1918, I consulted Henry Head, an eminent psychologist, and known for wise hunches as a doctor. The traditional prescription was complete rest, but he told me to work as soon as I felt like it (I had leave of absence) and as much as I felt up to, *but* — only in the morning. After a month or two I discovered, that, for me at least, morning work was far the better. I now never work after 6:30 p.m.

WARMING UP Most people need half an hour or so before being able to concentrate fully. I once came across some wise advice on this, and have taken it. The natural impulse towards the end of a day's work is to finish the immediate job: this is of course right if stopping would mean doing work all over again. But try to end in the middle of something; in a job of writing out, stop in the middle of a sentence. The usual recipe for warming-up is to run over the latter part of the previous day's work; this dodge is a further improvement.

Before coming to the subject of holidays I will say something about the various symptoms of overwork. I have wrongly disregarded them in the past; so doubtless others do too. One symptom can be muscular trouble. I once got into a vicious circle of feet painful enough to prevent exercise. I went to a masseuse who had the reputation of being a bit of a crank outside her work; she said my trouble was due to mental overwork. I am afraid I laughed, but I found she was perfectly right.

An ominous symptom is an obsession with the *importance* of work, and filling every moment to that end. The most infallible symptom is the anxiety dream. One struggles tensely all night with a pseudo-problem — possibly with some odd relation to one's current job — and wakes in the morning quite unrefreshed.

HOLIDAYS A governing principle is that 3 weeks, exactly 21 days — the period is curiously precise — is enough for recovery from the severest mental fatigue provided there is nothing actually pathological. This is expert opinion and my experience agrees entirely, even to the point that, e.g., 19 is not enough. Further, 3 weeks is more or less essential for much milder fatigue. So the prescription is 3-weeks holiday at the beginning of each vacation. It is vital, however, that it should be absolutely unbroken, whatever the temptation or provocation.

I believe there is a seasonal effect in mental activity, the trough coming about the middle or end of March. The academic year is surely involved, but there is probably a climatic factor. If this question is to be studied — and I think it should be — moments of high inspiration are far too rare and sporadic for statistics, but we could fall back on chess, bridge, and particularly crosswords, as tests of alertness and minor creation. My own evidence comes from a period of about 10 years when I played a form of two-pack patience, or solitaire, involving a very high element of skill. My curve of successes showed a general rise, with an unmistakable seasonal effect, a trough at the end of March. There was one anomaly, a March with a positive peak. I then recalled that I had taken a sabbatical term's leave from January through March, skiing in Switzerland.

HOLIDAY ACTIVITIES For many people these are highbrow: visits to Italian art galleries, a tour of Greece, and so on. I admire them, but do not share

their tastes. I have an intense interest in music, but this does not need travel. I began skiing at 40 and rock-climbing at 43: these activities renewed my youth. They give the most complete relaxation possible for mental work, and I believe they enlarge one's personality. Simple walking in mountains is very good. So is golf.

FOOD, TOBACCO, ALCOHOL Do not work within two hours of a substantial meal; blood cannot be in two places at once. I was once trapped into cold salmon at 6:30 p.m., immediately followed by a lecture, which I had had to leave largely to the spur of the moment. The lecture was confused and I was poisoned for a week afterwards. I should have starved.

On tobacco my bleak advice is: no smoking till the day's work is over. There is much to be said against regular smoking: you are merely normal when smoking and miserable when not. I was converted from the heaviest possible smoking (16 pipes and 4 cigars

a day) by Henry Head. He had the 1918 flu, followed by lack of tobacco on his ship and at his destination; an interval of 4 or 5 weeks. He decided to try giving it up for good. He then found that a heavy paper he was writing was finished in, he said, a third of the time it would have taken before. I sighed, succeeded in the struggle for abstinence, and I fully agreed with his estimate. I said speed was important; smoking is its enemy.

Alcohol is a depressant or sedative, not a stimulant, in spite of the illusion of champagne. It has a very valuable function in *stopping* thinking at the end of the day's work, a thing which many workers find indispensable. It has been said that, for this reason, Beethoven's posthumous quartets were paid for by cirrhosis of the liver. A permissible use is to mitigate the boredom of long routine calculations, or making a final draft. But then there must be a final check.

TEA AND COFFEE These are admirable, with no later reaction. They are usually considered to be rather mild in the way of stimulants; but since most of us never have the experience of being without them, it is possible that their virtue is underestimated.

Routine Chores

It is a good idea to keep a set period in the week to deal with these. As a young man I found it almost impossible to answer letters, so perhaps this is true of others. It is fatal (as I found) to put them into a file marked "Urgent." The technique I recommend is: will this letter ultimately be answered? If yes, answer immediately or in the set period. If no, straight into the wastepaper basket and forget it. This needs *knowing* oneself, a very important thing in many ways: not everyone does. There is a true story of an English clergyman offered a Colonial Bishopric. An inquiring visitor was told by the daughter of the house: "Father is in the study praying for guidance, Mother is upstairs packing."

DRUGS I can envisage a future in which stimulant drugs could raise mental activity for a set period of work, and relaxing ones give a suitable compensating period, perhaps of actual sleep. The present is a time of transition; stimulants do exist; but they should be used only with the greatest care and only in a crisis. And there is the problem of knowing what *is* a crisis.

. . . stimulants do exist; but they should be used . . . only in a crisis. And there is the problem of knowing what is a crisis.

From the Journal of the Association for Computing Machinery, January 1972:

3.2. THE GAP THEOREM. Informally, Theorem 3.6 guarantees that "in certain complexity ranges, we need only increase the computing power by some fixed function in order to achieve a genuine increase in computing capability." In contrast, Theorem 3.7 will show that "there exists arbitrarily large gaps in some complexity ranges where no new computation is performed." The following theorem was first proven by Trachtenbrot [23] and independently by this author.

THEOREM 3.7. (Gap Theorem). *Let* Φ *be a complexity measure, g a nondecreasing recursive function such that* $(\forall x)\, g(x) \geq x$. *Then* $\exists$ *an increasing recursive function t such that*

*Further to this interesting theorem, see page 120

Solutions to Puzzles by Smullyan

§1. (a) You have merely to make the one true statement: "I am not a knight".

(b) You have merely to make the one false statement: "I am a knave".

(c) This is more tricky! You must take some proposition whose truth-value is unknown to the King — for example that you are now carrying exactly eleven dollars in your pocket. Than you could say: "Either I am a normal and am now carrying exactly eleven dollars in my pocket, or else I am a knave".

Neither a knight nor a knave could make that statement, and the statement is true if and only if you are carrying exactly eleven dollars.

(d) No amount of statements can convince the King that you are not normal, since anything you say a normal could also say, because a normal can say anything.

§2. A question which works is: "Is *Bal* the correct answer to the question whether the statement that you are human is equivalent to the statement that there is gold on the island?" If he answers "Bal", then there is gold on the island; if he answers "Da", then there isn't. I leave the proof to the reader.

§3. First we must prove the following lemma: For any listed set A, there is a number n such that P_n is a knight if and only if $n \epsilon A$. Well, take any listed set A. The set A* of all n having at least one associate in A is a listed set (by c_5), hence has some index h. Therefore, for all n, $n \epsilon A_h \leftrightarrow n$ has at least one associate in A. Therefore, $h \epsilon A_h$ if and only if h has an associate in A.

Suppose $h \epsilon A_h$. Then h has some associate k in A. By definition of *associate*, P_k claims that h is extraordinary. Well, h is extraordinary ($h \epsilon A_h$), so P_k is making a true statement, so he is a knight. So in this case, $k \epsilon A$ and P_k is a knight.

Suppose $h \notin A_h$. Then h has no associate in A, yet h does have some associate k. Then k is outside A. Now P_k claims that h is extraordinary, but now h is not extraordinary. Therefore P_k is a knave. So in this case, P_k is a knave and k is not in A. This proves the lemma.

Now, the set E of Gödel numbers of all the established knights is listed, hence so is its complement E′. Then by the lemma, there is a number k such that P_k is a knight iff $k \epsilon E′$. But $k \epsilon E′$ iff P_k is not an established knight. Therefore P_k is a knight iff he is not an established knight. This means that he is either a knight but not an established one, or a knave who is an established knight. The latter alternative is impossible, so he is an unestablished knight.

As for the existence of an unestablished knave (which, incidentally, does not require the complementation condition c_3), apply the lemma to the set F of Godel numbers of all the established knaves. Then there is a number n such that P_n is a knight iff $n \epsilon F$, so P_n is a knight iff P_n is an established knave. Since P_n cannot be both a knight and an established knave, then he can't be either one. Hence he is an unestablished knave.

(b) Suppose the set K of Gödel numbers of all the knights were a listed set. Then its complement K′ would be listed. Then by the lemma, there would be a number k such that P_k is a knight iff $k \epsilon K′$. But $k \epsilon K′$ iff P_k is not a knight. Therefore P_k would be a knight if and only if he isn't, which is a contradiction. Therefore the set K cannot be listed.

§4. Take any bird A. By C_2 (taking M for B) there is a bird B_h (called "C" in C_2) such that for every n, $A(M(n)) = B_h(n)$. This means that for every n, $A(B_n(n)) = B_h(n)$ (since $B_n(n) = M(n)$). Taking h for n, we have $A(B_h(h)) = B_h(h)$. So A is fond of the number $B_h(h)$.

This proves that every bird is fond of at least one number. Hence also the mocking bird M is fond of some number k -- M(k) = k. This means that $B_k(k) = k$, so the bird B_k is egocentric.

§5. For any flowers A, B, let $A \downarrow B$ be the flower which is blue on just those days when A, B are both red (there cannot be more than one such flower by c_1). Let A′ be the flower $A \downarrow A$. Then A′ always has different color from A. Let $A \cap B$ be the flower $(A \downarrow B)'$. Then $A \cap B$ is blue on just those days when A, B are

both blue. Let $A \cup B = (A \cap B)'$, and $A \cup B$ is blue on just those days when at least one of A, B is blue. The set of flowers forms a Boolean algebra with respect to the operations $\cap$, $\cup$, $'$ and it is a well known fact that the number of elements of any finite Boolean algebra must be a power of 2. Therefore (by C_3) there are exactly 256 flowers in the garden.

Solution to the Problem in Number 1, p. 58

Imagine the street as being circular and as having n + 1 spaces. Allow the drivers to choose the (n + 1) st space. (This means they can choose to fail.) The sample space then has $(n + 1)^n$ points. For any point of the sample space the cars can now proceed to park, and when they are done there will be just one empty space. The point corresponds to a success in the original problem if and only if the empty space is the new (n + 1) st space. The mapping from the $(n + 1)^n$ - point sample space to the n + 1 possible empty spaces has the same number of points in each preimage. Therefore this number is $(n + 1)^n / (n + 1) = (n + 1)^{n-1}$, as was to be shown. *(H. Edwards)*

A Problem

The host at a dinner party has invited an unequal number of men and women, and he wants to seat the guests at a circular table in such a way that the sexes are evenly distributed. He will regard them as evenly distributed if the following criterion is met: For any positive integer k and for any two sets of k consecutive places around the table, the numbers of men and women in the two sets are either identical or differ by just one. Show that it is always possible to find a pattern for seating the men and women so that they will be evenly distributed. Show, moreover, that, up to rotations, there is only one such pattern. *H.E.*

```
      W M W
   W        M
 M            W
W              W
W              M
 M  W        W
   W M W
```

Even distribution of 6 men and 10 women.

---

* **Corrigendum**

There was an omission in the printing of Theorem 3.7, the Gap Theorem, page 166. It should read:
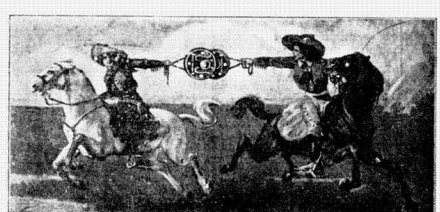
THEOREM 3.7. (Gap Theorem). *Let Φ be a complexity measure, g a nondecreasing recursive function such that $(\forall x)g(x) \geq x$. Then $\exists$ an increasing recursive function t such that no Φ_i satisfies $t(x) \leq \Phi_i(x) \leq g \circ t(x)$ i.o. x.*
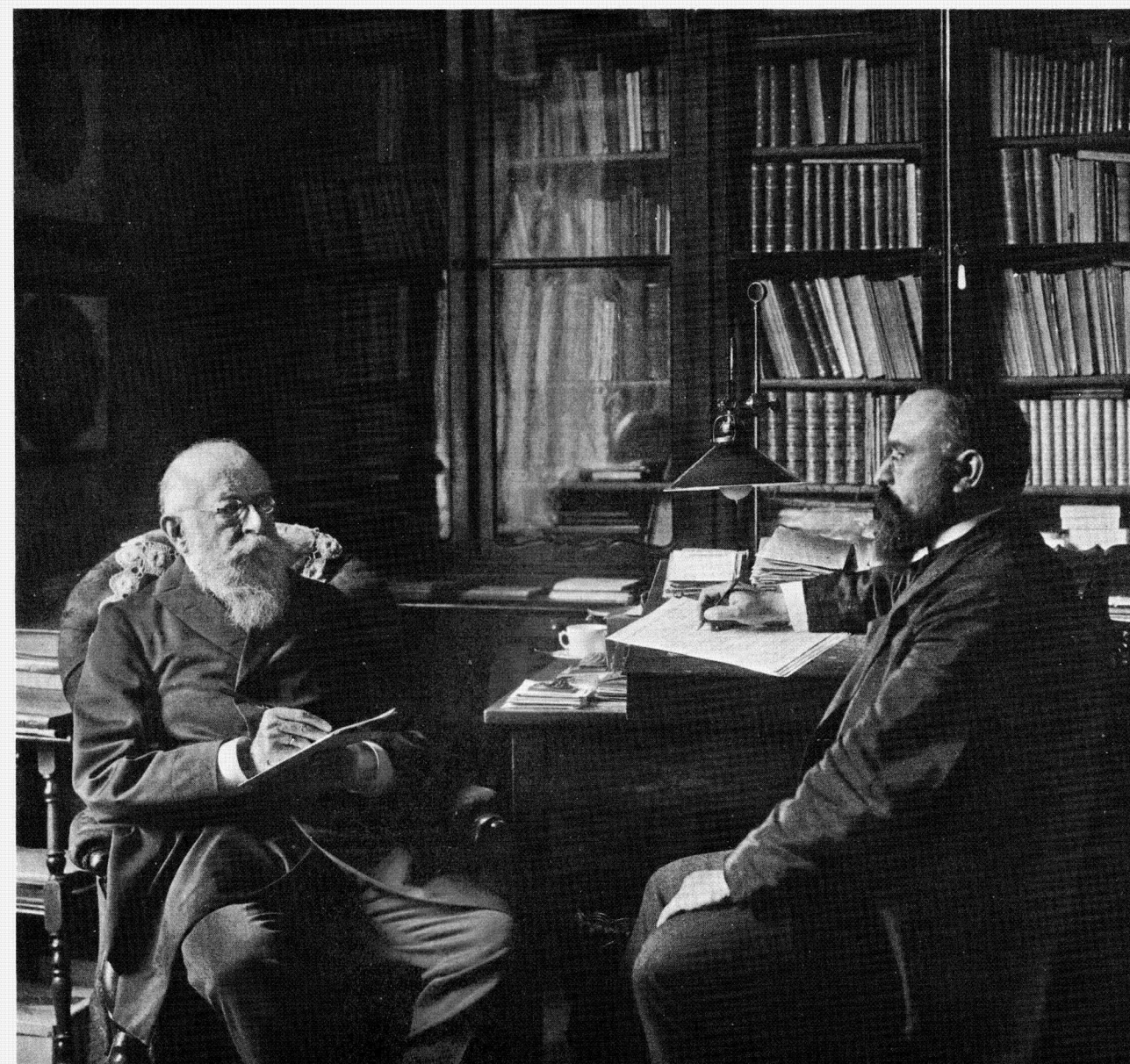
F. Prÿm   *G. Rost*

The reference to Friedrich Emil Prÿm in Erwin Neuenschwander's article "Riemann's Example of a Continuous Nondifferentiable Function" (Mathematical Intelligencer Volume 1 No. 1 pp 40-44) stirred some dormant recollection that F.E. Prÿm not only developed the theory of generalized Theta functions and, together with Georg Rost, a concept of general Abelian functions later called Prymsche Functionen but that he also invented the spring press dress fastener (snaps, to Americans) as advertised by Rylands and Sons in 1913. At the risk of debunking an attractive myth, the Intelligencer feels obliged to reveal the truth: the black sheep of the family, F.E. Prÿm devoted himself largely to mathematics and played little or no role whatsoever in the family business whose factory manufactured these press fasteners.

Actually, this short note was really prompted by the "discovery" of the truly impressive picture of Prÿm and Rost at work which is reproduced from F. Prÿm and G. Rost "Theorie der Prymschen Functionen Erste Ordnung", B.G. Teubner, Leipzig, 1911.